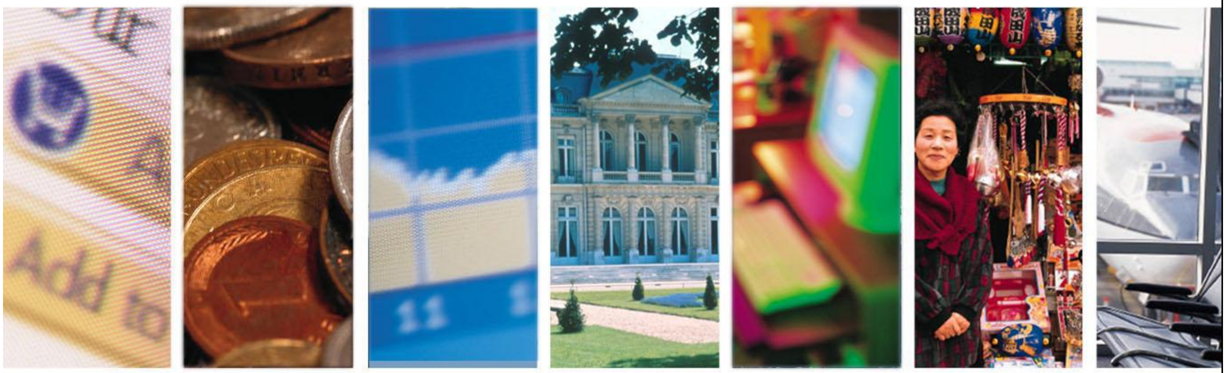
ORGANISATION FOR ECONOMIC CO-OPERATION AND DEVELOPMENT



FORUM ON TAX ADMINISTRATION: SMALL/MEDIUM ENTERPRISE (SME) COMPLIANCE SUB-GROUP

Guidance Note

Evaluating the effectiveness of compliance risk treatment strategies

November 2010

CENTRE FOR TAX POLICY AND ADMINISTRATION

Table of contents

Figures

Boxes

ABOUT THIS DOCUMENT

Purpose

Building on the Forum's prior work on compliance risk management, this note provides practical guidance on a structured and systematic approach that can be used by revenue bodies, to the extent deemed practicable, to evaluate the effectiveness and impacts of their compliance risk treatment strategies.

Background to the Forum on Tax Administration

The Forum on Tax Administration (FTA) was created by the Committee on Fiscal Affairs (CFA) in July 2002. Since then the FTA has grown to become a unique forum on tax administration for the heads of revenue bodies and their teams from OECD and selected non-OECD countries.

In 2009, participating countries developed the *FTA vision* setting out that… *The FTA vision is to create a forum through which tax administrators can identify, discuss and influence relevant global trends and develop new ideas to enhance tax administration around the world***.** This vision is underpinned by the FTA's key aim which is to…*improve taxpayer services and tax compliance – by helping revenue bodies increase the efficiency, effectiveness and fairness of tax administration and reduce the costs of compliance.*

To help carry out its mandate, the FTA is directly supported by two specialist sub-groups—Compliance and Taxpayer Services—that each carry out a program of work agreed by members. Both OECD and selected non-OECD countries participate in the work of the FTA and its sub-groups.

The Compliance Sub-group's mandate, in broad terms, is to provide a forum for members to:

- periodically monitor and report on trends in compliance approaches, strategies and activities;
- consider and compare members' compliance objectives, the strategies to achieve those objectives and the underlying behavioural compliance models and assumptions being used;
- consider and compare member compliance structures, systems and management, and staff skills and training; and
- develop and maintain papers describing good country practices as well as develop discussion papers on emerging trends and innovative approaches.

Since its inception, the Sub-group has focused its work on issues associated with improving the tax compliance of small to medium enterprise (SME) taxpayers.

Caveat

National revenue bodies face a varied environment within which to administer their taxation system. Jurisdictions differ in respect of their policy and legislative environment and their administrative practices and culture. As such, a standard approach to tax administration may be neither practicable nor desirable in a particular instance. The documents forming the OECD tax guidance series need to be interpreted with this in mind. Care should always be taken when considering a country's practices to fully appreciate the complex factors that have shaped a particular approach.

Inquiries and further information

Inquiries concerning any matters raised in this guidance note should be directed to Sean Moriarty (International Co-operation and Tax Administration Division) at e-mail (Sean. Moriarty@oecd.org).

Summary

Revenue bodies have a responsibility to use their limited resources in the most cost effective manner. In a tax compliance context, exercising this responsibility requires them to have a systematic approach for managing tax compliance risks, aiming to ensure that such risks are properly identified, assessed, prioritised and treated. Given the cyclical nature of this process, a key requirement is that revenue bodies are able to understand the impacts achieved and effectiveness of their risk treatment strategies. For many strategies, this is not an easy task.

This guidance note sets out a practical methodology for conducting outcome evaluations of compliance risk treatment strategies in priority areas. It draws on innovative work, based on considerable research, carried out by one OECD revenue body, and considerable further assistance from a number of other revenue bodies that have been intensifying their efforts to better understand the impacts/effectiveness of their compliance program activities. After outlining a number of important concepts and issues that are often raised in an evaluation context (e.g. output/ outcome program model and attribution), the guidance note introduces and elaborates a four phase compliance effectiveness methodology:

- Phase 1 is about understanding and articulating the risk and making sure it aligns with the revenue body's business intent.

- Phase 2 is about clearly expressing a revenue body's desired outcomes from its compliance strategies and what would be different if it was successful. It also involves identifying the right mix of strategies which treat the drivers of compliance behaviour and not just the behaviour.

- Phase 3 involves identifying potential indicators for each compliance strategy and validating them to ensure they are viable and useful, and capable of providing a credible picture of the effectiveness of a revenue body's compliance strategies.

- Phase 4 deals with evaluating and reporting on the extent to which a revenue body has been effective in changing compliance behaviour, building community confidence, or both, over the immediate, intermediate and long term.

The guidance note gives emphasis to the identification and development of practical measures and indicators that, set at the commencement of risk treatment action, can be used to gauge progress and assess the merits of risk treatment strategies in terms of improved taxpayer compliance and confidence in a revenue body's administration. To illustrate these aspects, the note includes a large number of case study examples from a cross-section of revenue bodies, describing specific measures and indicators used to evaluate the impacts of specific risk treatments and, for most, the impacts observed. The note acknowledges that for some strategies there are limitations as to degree of precision/confidence that can be attached to conclusions concerning their impacts and effectiveness.

Finally, the note acknowledges that it is not possible, nor even desirable, for a revenue body to fully evaluate every specific risk treatment carried out, given the burden this would present. In this regard, it emphasises the need for revenue bodies to adopt a pragmatic approach in deciding what is important in an evaluation context and provides practical guidance to assist in exercising such judgments. The guidance note is accompanied by two other documents that provide important reference information to its use—a background note '*Selected revenue body and other experience with implementing risk treatment evaluation*' and a companion guide '*Methodological techniques for use in evaluating the effectiveness*'.

Recommendations

- Revenue bodies are strongly encouraged to examine the evaluation guidance provided and to consider its relevance and practical application in their respective contexts;

- Revenue bodies are strongly encouraged to have in place a sound and practical approach to evaluation that can be relied to provide credible evidence pointing to the effectiveness of their key compliance risk treatment strategies.

I. Introduction

Background and context

1. In September 2004, the Committee on Fiscal Affairs approved publication of the guidance note titled '*Compliance Risk Management: Managing and Improving Tax Compliance'* (hereafter referred to as 'the 2004 guidance note') which provided a framework for the application of modern compliance risk management principles to the management of tax compliance risks. The 2004 guidance note also described a step by step strategic process for the identification and treatment of compliance risks, supported by a number of practical examples drawn from revenue bodies that served to illustrate particular approaches and their impacts. In doing so, it identified and discussed general principles found in both the identification and treatment of compliance risks within a variety of taxation jurisdictions and also provided information about the way in which treatment strategies influence the behaviour of small businesses in relation to their tax obligations. To provide a context for this guidance note, it is useful to recap the key themes from the 2004 guidance note.

Domains of taxpayer obligation

2. The 2004 guidance note considered that while specific obligations of taxpayers will vary from one taxation role to another and from one jurisdiction to the next, four universal categories of obligation are likely to exist for almost all taxpayers. These broad categories of taxpayer obligation are:

   1) registration in the tax system;
   2) timely filing of tax returns;
   3) reporting of complete and accurate information in tax returns; and
   4) payment of tax obligations on time.

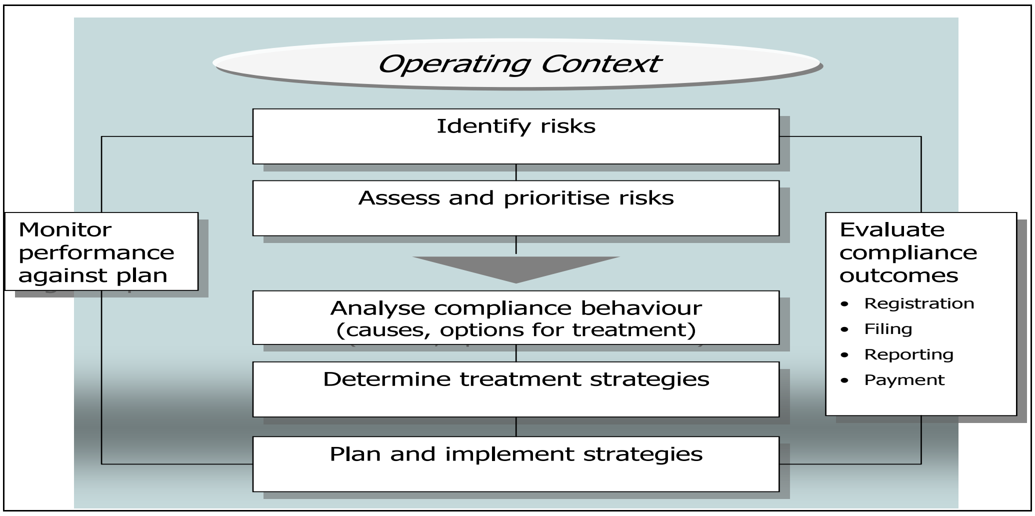
   It found that, irrespective of jurisdiction, "compliance" will relate to the extent to which taxpayers meet these obligations. Taxpayers who fail to meet any of them may be considered to be "non-compliant", although the reason for their non-compliance may fall along a continuum extending from unintentional error to intentional/deliberate actions.

Adopting a risk management approach

3. The 2004 guidance note acknowledged that revenue bodies operate with finite resources, and that effective tax administration needed to focus on optimising collections under the tax laws in ways that instil confidence in a revenue body and ensured the tax system is operating and being seen to operate correctly. It promoted the notion that revenue bodies needed a structured and systematic process for deciding what is important in a tax compliance context, how major compliance risks will be addressed and how to determine the most effective allocation of their resources — the compliance risk management process.

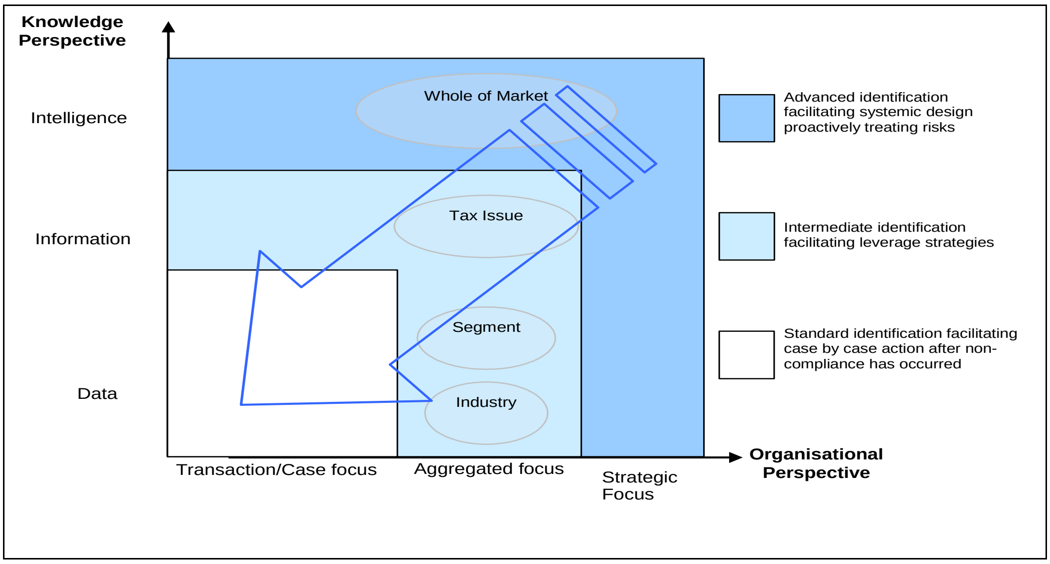
4. The 2004 guidance note also outlined compliance risk management as a series of defined (yet iterative) steps for the systematic identification, assessment, ranking, and treatment of tax compliance risks, as well as related monitoring and evaluation activities, to support improved decision-making. A model of the compliance risk management process for use by revenue bodies was presented (see Figure 1). Through greater understanding of the environmental context, this model would enable revenue bodies to assess and determine which risks/ taxpayer behaviours it can or cannot influence, and to consider and prioritise which risk treatments were required.

Figure 1. The compliance risk management process



5. In practice, compliance risks fall on a continuum and therefore an additional diagnostic model (see Figure 2) was used to illustrate that as the level at which risks are managed increases, there is a need for a commensurate increase in the level of evidence and knowledge that both informs and substantiates the strategic risks identified for treatment.

Figure 2. The compliance risk management diagnostic model



Recognising the influences on taxpayers' compliance behaviour

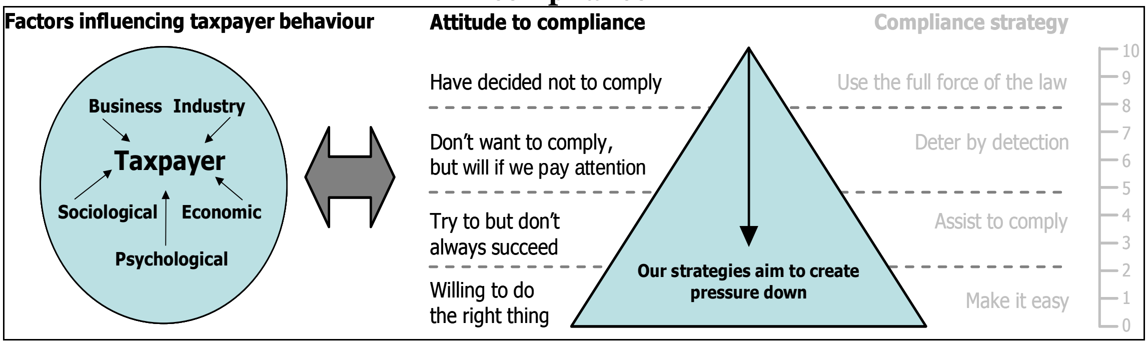
6. Research supporting the 2004 guidance note described two broad approaches to the problem of compliance, the first based on economic rationality and the second based on wider behavioural issues [1]. It considered that these approaches were not necessarily competing and that each approach can be valuable in terms of understanding tax compliance.

7. In combination with the research of Australian academic Dr Valerie Braithwaite around the sets of values, beliefs and attitudes of individual taxpayers (persons or businesses),

[1] James, Simon, Hasseldine, John, White, Peggy and Toumi, Marika (2001) 'Developing a Tax Compliance Strategy for Revenue Services', *Bulletin from International Bureau of Fiscal Documentation*, April 2001, pp. 158-164.

the 2004 guidance note offered a BISEP model for thinking about the combination of these factors and the attitudes of individuals in the way they may relate to a revenue body and the tax system it administers. This model (see Figure 3) allows users to categorise the factors in a consistent manner and form to aid understanding and is also sufficiently robust to stand up to applied and practical application in a tax administration environment.

Figure 3. The BISEP model & spectrum of taxpayer attitudes to compliance



8. It was highlighted that an individual taxpayer is capable of adopting any of the attitudes described at different times and may also adopt all of the attitudes simultaneously in relation to different issues. It noted that these attitudes are not fixed characteristics of a person or group, but reflect the interaction between the person or group and those that impose demands upon them.

9. The 2004 guidance note considered that the value of this model is in the contribution it makes in developing a deeper understanding of taxpayer behaviour and being able to lay the groundwork for the development of targeted strategies which encourage the motivation to do the right thing and constrain the motivation to resist or evade compliance. Today, many revenue bodies (e.g. Denmark, New Zealand and UK) have adopted this model, or a close variant of it, to guide their planning for delivering improved compliance outcomes.

Monitoring and evaluation of compliance risk management strategies

10. A key element of the recommended compliance risk management process was a compliance measurement framework that would provide revenue bodies with a range of compliance indicators that could be used to monitor and evaluate the impacts of their compliance activities, both at the aggregate/macro level and in respect of specific risk treatments strategies. Such a framework would enable a continual cycle of review and refinement. The 2004 guidance note did not, however, provide detailed guidance on likely sets of compliance indicators to be used or the steps to be followed for implementing such arrangements, given this was still a largely unexplored area of tax administration. This shortcoming was partially addressed in the 2008 guidance note '*Monitoring Taxpayers' Compliance: A Practical Guide Based on Revenue Body Experience*' which explored and promoted the idea of a compliance monitoring framework at the aggregate/macro level.

Benefits of adopting this approach

11. The 2004 guidance note considered a revenue body could derive a range of benefits by adopting the compliance risk management approach:

- It would provide a systematic and structured basis for strategic planning;
- By focusing on the underlying drivers (not symptoms) of non-compliance, it gave greater flexibility for customising and tailoring solutions to specific compliance risks, rather than adopting a 'vanilla' approach;

- Better outcomes in terms of both effectiveness (increases in compliance with tax laws) and efficiency (cost to administer) could be anticipated;
- There would be increased organisational agility developed through evidence-based evaluation, continuous improvement and learning; and
- It would improve the transparency of its approach to managing compliance and enhance its ability to withstand external scrutiny.

This guidance note

12. At meetings of the Compliance Sub-group in both 2008 and 2009, members acknowledged that there was a critical gap in the detailed practical guidance available to revenue bodies for fully implementing the recommended risk management process. The subject of this gap was the practical approaches and methods that could be used for systematically evaluating the impacts of specific compliance risk treatment strategies. However, it was acknowledged that work was underway by both the Australian Taxation Office (ATO)[2] and by European Commission (EC) (as part of its *Fiscalis* program [3]) to develop practical guidance in this area which might serve as valuable input to guidance that the sub-group could prepare. Members accordingly agreed to initiate work to develop a set of practical guidance in this field.

13. In giving direction for this work members requested that the guidance should build on work already done in this field, limited as it is, and not aim for absolute precision, recognising that evaluation in the field of taxpayers' compliance was 'more of an art than a science'. Furthermore, it should encompass ideas for its practical implementation in an organisational sense and, in particular, should:

    1) be oriented towards senior managers (as opposed to technicians);
    2) be practical and not too academic;
    3) have a clear 'outcomes' orientation;
    4) provide an overview of measurement approaches that are feasible; and
    5) be supported by good case study examples to demonstrate the recommended techniques.

14. The guidance provided in this note, along with background information contained in the accompanying documents— a background note '*Selected revenue body and other experience with implementing risk treatment evaluation*' and a companion guide '*Methodological techniques for use in evaluating the effectiveness*'—draws on the formal compliance effectiveness methodology and related practices developed by the ATO and Forum members are indebted to the ATO for its initiative in this field of revenue administration. Documentation prepared as part of the European Commission's '*Overview of Evaluation Methodology and Techniques*' has also provided valuable insights and reference materials. Considerable further assistance, including an extensive series of related case studies, was also received from revenue bodies in Australia, Canada, Denmark, New Zealand, Singapore, Sweden and the United Kingdom in the planning and preparation of this guidance note.

---

[2] All of the ATO published guidance and research on methodologies for measuring the effectiveness of its compliance strategies is described in Annex 2.

[3] The European Commission's Fiscalis program has prepared practical guidance on risk management, including evaluation, as part of a Compliance Risk Management Guide for Tax Administrations, a revised edition of which was published in March 2010 —see http://ec.europa.eu/taxation_customs/resources/documents/common/publications/info_docs/taxation/risk_managt_guide_en.pdf

II. Key concepts and issues in monitoring taxpayers' compliance and evaluating compliance risk treatment strategies

15. This chapter provides background to a number of important concepts and issues that are often raised when discussing tax compliance issues in general and compliance monitoring and evaluation in particular. They are elaborated here to inform readers and to serve as a foundation for exploration of the issues dealt with in this note.

Measuring revenue body performance—outcomes and outputs; effectiveness and efficiency

16. The notion of monitoring taxpayers' compliance is directly relevant to the terminology of 'outcomes' and 'outputs' and 'effectiveness' and 'efficiency' when used in relation to measuring the performance of revenue bodies, both at a macro/whole of tax level, or in respect of specific compliance strategies, as dealt with in this guidance note. As these terms are often confused and sometimes used interchangeably, it is considered helpful to provide some clarification in a tax compliance context.

17. There is a considerable body of literature on performance measurement in general and government oversight bodies in many countries have formalised their definitions as part of agency guidance. The key points are as set out in the comments that follow.

Outcomes and outputs

18. The term '*outcomes*' is typically associated with 'impacts', 'results', and 'external consequences'. In a tax compliance context, this can clearly be related to 'levels of voluntary compliance' (e.g. in relation to timely return filing, correct reporting of liabilities and timely payment of taxes), noting that self assessment regimes operate on the fundamental principle of voluntary compliance. Also relevant in an 'outcomes' context are changes in taxpayers attitudes to, and perceptions of, tax compliance, and taxpayers' degree of confidence in the administration by the revenue body, as these are both clearly influenced by the work (and effectiveness) of revenue bodies.

19. Measuring outcomes also has an important time dimension. In practice, a compliance strategy may entail a number of risk treatments that impact in different ways and over different time periods, and which may lead to immediate, medium term and/or long term outcomes. This observation is particularly pertinent in relation to revenue bodies whose management and operations are heavily geared to a 'fiscal year cycle/ orientation'. However, for determining the impacts of a compliance strategy, the evaluation will very often need to have a short, medium and longer term horizon, in order to gauge the extent to which expected outcomes are being achieved and to ascertain if these are sustained over time.

20. '*Outputs*', on the other hand, typically refer to the 'products' of internal activities undertaken by an organization within a defined period (and generally covering no longer than a year) to produce the outcomes sought. For a revenue body this would include tax revenue, answered inquiries, completed audits and finalised debt cases.

Effectiveness and efficiency

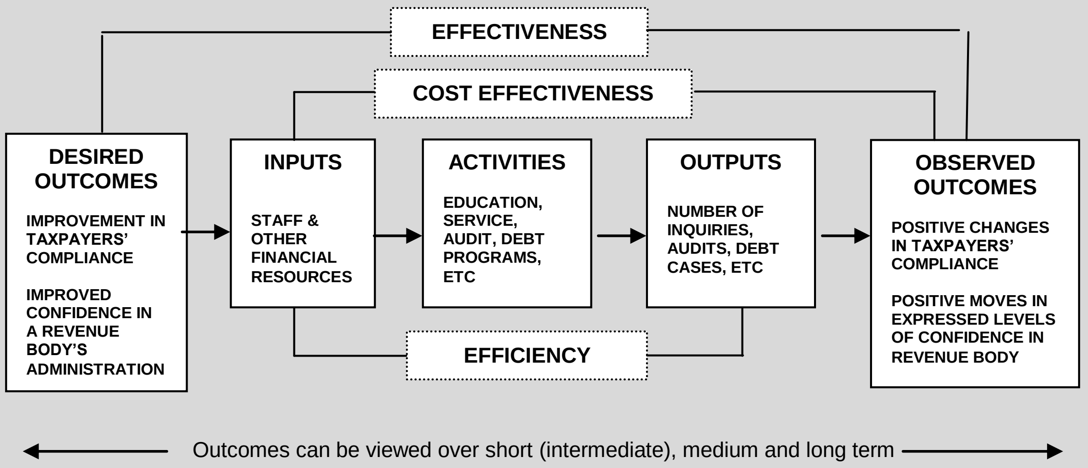
21. The term '*effectiveness*' is typically associated with the extent to which improved '*outcomes*' are being achieved. In a tax compliance context, the extent to which compliance (e.g. filing, reporting and payment) has been improved as a result of a revenue body's activities would clearly be an indication of its increasing effectiveness (acknowledging, of course, that in some areas its measurement may be difficult to gauge in precise terms, as well as the issue of attribution which is examined later in this note. Similarly, a trend of more positive attitudes to, and perceptions of, tax compliance and/ or improved confidence in a revenue body's administration of the tax laws could be seen as indicators of its increasing effectiveness.

22. '*Efficiency*' relates to reducing or minimising the use of resources to produce a given level of outputs (e.g. increasing the number of completed audits for a given level of staffing, other things being equal, would reflect improved efficiency) or conversely, increasing the volume of outputs for a given level of inputs.

23. The relationships between these terms in a performance measurement context are often presented by way of a 'program logic' diagram or chart. An example of this in a tax compliance context is provided in Figure 4 [4]. Applying the model, a program typically entails the following steps: 1) clarify the objectives (i.e. what outcomes are being sought); 2) map the connections between the inputs, activities, outputs and outcomes; 3) identify the level of outcomes to be measured (both intermediate and final); 4) defining how success will look; and 5) defining what performance information will be used.

24. Within this model, efficiency measures (e.g. number of audits performed per staff resource) reflect the relationship between outputs and the inputs used to produce them, while effectiveness measures reflect the outcomes achieved (e.g. extent of change in compliance achieved) vis-à-vis the desired outcomes/objectives (e.g. extent of change in compliance sought).

Figure 4: Tax compliance program logic



25. As indicated earlier, revenue bodies have traditionally accounted for their performance (including for their compliance improvement activities) largely (and for some, solely) in terms of the 'outputs' resulting from the various work streams administered by them (e.g. revenue collection, audits and debt collection). However, while output measures represent an important part of revenue bodies' reporting systems and assist in day to day management, they do not provide any insights as to the outcomes or impacts resulting from their activities. A related concern is that positive trends in output volumes (e.g. total tax revenues and additional tax assessed resulting from audit action) are sometimes inferred (without adequate accompanying evidence) as reflecting improvements in taxpayers' compliance, but this need not necessarily be the case (e.g. over-budget revenue collections may be the result of estimates based on poor data and assumptions, significant increases in tax assessed from audit programs may result from a few abnormally large audit cases). These sorts of considerations suggest that a degree of caution should be applied when inferring improved compliance from what appear to be positive trends in output-related measures.

[4] As referenced in '*Literature Review—Measuring Compliance Effectiveness*' (ATO) May 2007.

Measures and indicators

26. Having introduced and defined the terms '*effectiveness*' and *'outcomes'* in a tax compliance context, it is appropriate to discuss how they are measured. For this purpose, it is relevant to introduce the terms '*measures*' and '*indicators*' as these are used widely in the rest of this note.

27. An OECD report (unpublished) prepared in 2001 [5] made a clear distinction between 'measures' and 'indicators'. *'Measures'* correspond to the expected direct results at any program level—they are a quantification of the actual program results achieved (e.g. the number of taxpayers who file returns on time). On the other hand, '*indicators*' are less direct measures, used where direct measures are impractical or unavailable; an example is a decline in the proportion of productive audits being used as an indication of improved taxpayers' compliance. Generally speaking, an indicator does not provide irrefutable evidence of the result being inferred and, in the absence of other indicators, should generally be used as a pointer for further inquiry [6]. On the other hand, a range of indicators all pointing to a similar conclusion may provide a stronger or more credible case for the conclusion/result being argued.

Attribution—Understanding the factors that influence taxpayers' compliance and behaviours

28. In brief, attribution is about understanding the relationship between 'cause' and 'effect'. It has been defined as.... "the extent to which a reasonable causal connection can be made between a specific outcome and the activities and outputs of a government policy, program or initiative" [7]. For reasons explained hereunder and elsewhere in this note, this is often not a straightforward and easy task to execute.

29. The 2004 guidance note (see page 36 et seq.) emphasised the range and nature of factors that can influence tax compliance and taxpayers' behaviour. These include economic and social conditions, involving 1) national and global economic factors; 2) physical environment factors; and 3) legislative, social, economic, demographic attitudinal factors. From a tax compliance monitoring viewpoint, the existence of such factors may complicate determining the extent to which changes in observed behaviour can be attributed to the actions of the revenue body, as explained as follows. [8]

    "The attribution problem concerns how to determine whether a particular program caused the observed results, and to what extent.................. There are three main ways in which observed changes might not be attributable to compliance activities.

    Firstly, a change might have occurred regardless of a compliance activity or program and may be due to external or internal factors. Secondly, while a program may have been necessary for the result to have occurred, it may not have been sufficient for the result. That is, particular circumstances were also required for that program to have had the results it did. This is important if the results are to be generalised to other contexts, such as different places, times or situations. This is called external validity. Where neither the program nor the circumstances can be perfectly replicated, the inferences could be weak. This has particular relevance for decision making in an organisation. It is therefore advisable to use multiple evaluation strategies where there are significant threats to external validity (Treasury Board of Canada, 1998, pages 22.30). Thirdly, there may be no observable changes, but in fact a program

[5] *'Performance measurement in tax administrations'*, (OECD Centre for Tax Policy and Administration) 2001.

[6] The Canada Revenue Agency describes a compliance indicator as a quantitative or qualitative factor that provides a valid proxy measure of compliance achievement or change (Compliance Measurement Framework, CRA, September 2003).

[7] Treasury Board of Canada, '*Results-Based Management Lexicon*' 2004.

[8] See '*Literature Review—Measuring Compliance Effectiveness'* (Australian Taxation Office) May 2007.

may have halted deterioration and simply maintained the status quo. This does not mean that the program had no effect—it is still vital to know what would have happened had the program not been implemented."

30. It is also worth noting that outcome measures are often measured with a degree of uncertainty which further compounds the attribution problem.

31. More is said on the methods for inferring attribution in a tax compliance effectiveness context in the following chapter.

Direct and indirect effects of compliance risk treatments

32. Previous OECD materials have discussed the issue of the *direct* and *indirect* (ripple) effects of compliance treatments—see '*Compliance Measurement*' (2001) at pages 12-13, and the 2004 guidance note, pages 63-64.

33. The *direct* effects of compliance treatments refer to the results achieved with the targeted taxpayer population for the fiscal period under examination and in subsequent periods. *Indirect* effects, on the other hand, refer to the impacts of treatments among other taxpayers, both for the period under examination and for subsequent periods—the so-called "ripple effect" of compliance activities. *Indirect* effects may result directly from personal knowledge of the affairs of target taxpayers, and/ or indirectly through media publicity or some other medium of communication. By their nature, *indirect* effects are extremely difficult to isolate and quantify.

Benefits and costs of evaluating compliance risk treatment strategies

34. While revenue body experience with evaluation of specific risk treatment strategies is relatively limited it is possible to identify, in broad terms, the nature of the benefits and costs associated with its implementation.

35. Concerning benefits, the available revenue body experience and general literature [9] concerning evaluation point to the following:

    - improved knowledge of the impacts achieved from specific strategies and the extent to which they are producing the desired degree of improved outcomes;

    - improved knowledge of the efficacy of specific risk treatments, although this will be difficult for those strategies involving a fair mix of treatments (e.g. education, publicity, warning letters, audits) where it is not possible to differentiate the impacts of individual treatment actions;

    - promoting effectiveness, enhancing transparency/ accountability and facilitating learning through an increased focus on explicit assumptions and documentation relating to the risk, the choice of treatment and the outcomes being achieved;

    - the building of a credible evidence base to support future decision-making, and which may provide justification for changes in legislation or support for additional funding/ resources;

    - over the longer term, a more 'outcomes-focused' workforce; and

    - improved accountability through a capacity to provide evidence-based assurances that effective programs are being maintained and resources are not being wasted.

[9] For example, see page 5 of '*Introduction to program evaluation for public health programs: A self study guide*' (US Department of Health and Human Services (August 2005).

36. Implementing and administering evaluation does present a range of initial and ongoing investments/ costs that must be made by a revenue body. The more significant of these are set out hereunder:

    - executive and senior management time in leading and promoting the revenue body emphasis on risk treatment evaluation;
    - the establishment and operation of a central and/or decentralised evaluation capability (NB: the scale of this will vary according to a range of factors (e.g. size of organisation, the policy on evaluation priorities);
    - costs associated with acquiring new data and improving administrative data systems;
    - the time associated with training line management officials on new requirements to properly implement risk treatment evaluation;
    - the costs associated with producing and disseminating documentation/ publications concerning risk treatment evaluation.

37. Some further insights and as the nature and scale of these activities are set out in the accompanying background note '*Selected revenue body and other experience in implementing risk treatment evaluation*'.

III. Methodology for evaluating the impacts of compliance strategies

Understanding compliance effectiveness

38. Compliance effectiveness is the extent to which the actual outcomes of a revenue body's compliance strategies align with its desired outcomes—*Did the strategies employed have the effect or impact on compliance behaviour and/or community confidence that were expected of them*?

39. The compliance effectiveness methodology described in this note is a tool that can be used to plan the right mix of strategies, measure their effectiveness and identify opportunities for continuous improvement. A revenue body can use it for both planning and evaluation purposes. What it learns about its effectiveness may inform its future planning and help reshape its compliance improvement strategies.

40. There are four iterative phases in the recommended methodology (Figure 5 refers):

    - Phase 1 is about understanding and articulating the risk and making sure it aligns with the revenue body's business intent.

    - Phase 2 is about clearly expressing a revenue body's desired outcomes from its compliance strategies and what would be different if it was successful. It also involves identifying the right mix of strategies which treat the drivers of compliance behaviour and not just the behaviour.

    - Phase 3 involves identifying potential indicators for each compliance strategy and validating them to ensure they are viable and useful. The suite of indicators should be capable of providing a credible picture of the effectiveness of a revenue body's compliance strategies.

    - Phase 4 deals with evaluating and reporting on the extent to which a revenue body has been effective in changing compliance behaviour, building community confidence, or both, over the immediate, intermediate and long term.

41. Phases 1, 2 and 3 focus on planning to be effective while phase 4 focuses on evaluating the extent to which a revenue body has been effective.

42. When evaluating compliance effectiveness a revenue body needs to answer the following questions:

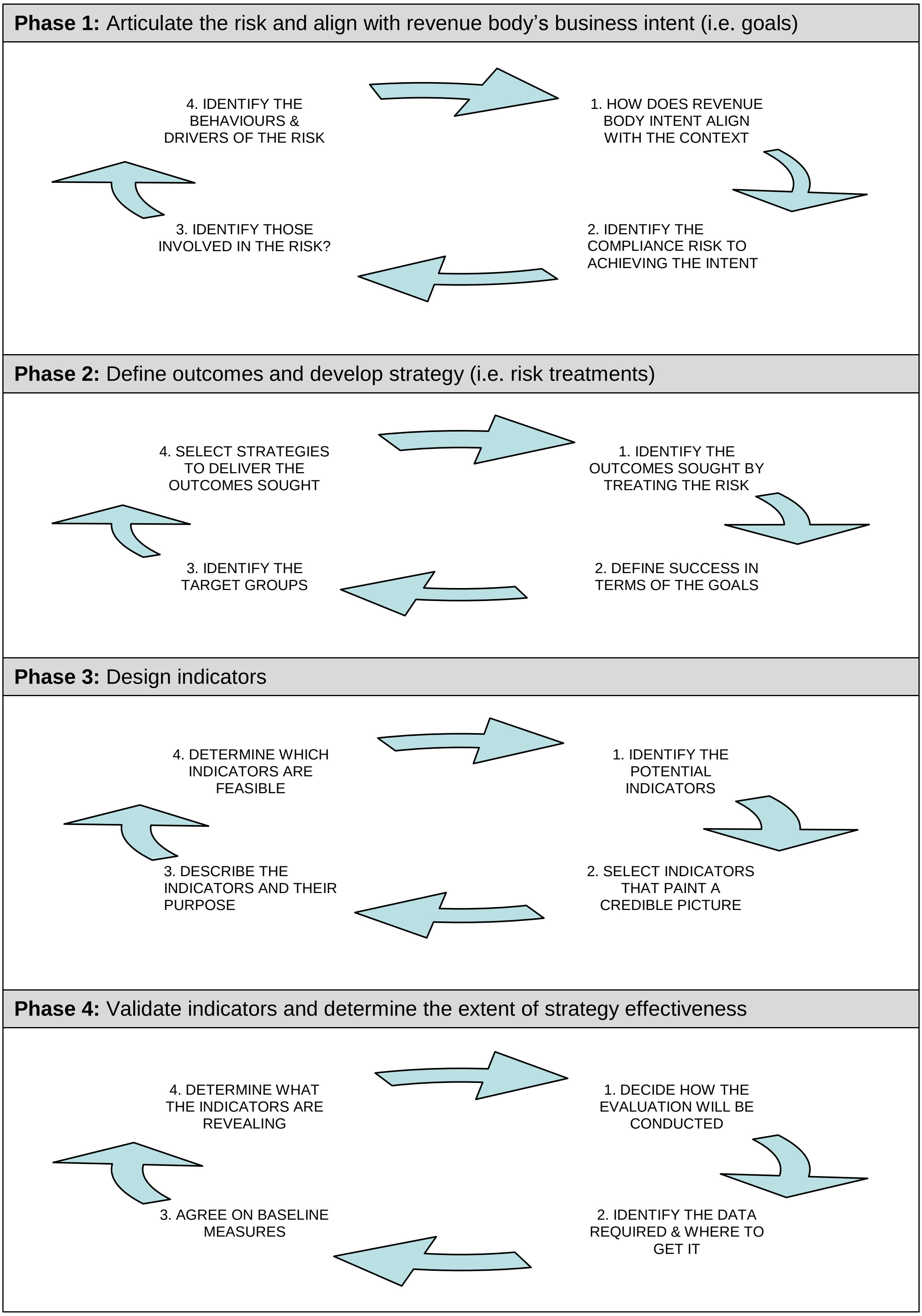
    - Is there a **change** in compliance behaviour and/or community confidence?
    - And if there is a change, did a revenue body's actions cause it (i.e. attribution)?

43. To understand this, a revenue body needs to explore behaviour at:

    - the broad or macro level [10] to identify any observable changes; and
    - the strategy or micro level to determine whether those changes can be attributed to its compliance strategies.

[10] The broad or macro level is a "high level" view that can bring to note year to year changes and, with enough data over time, trends in compliance in the general population. However, isolating all of the factors that may have influenced such changes, including those undertaken by a revenue body, is often quite difficult, and the process of attribution even more so. Many revenue bodies conduct regular surveys of the external community that gauge the population's perceptions and views on issues of interest (such as perceptions of service and compliance) over time that can reflect on the effectiveness of their administration.

Figure 5. Compliance effectiveness methodology



Setting the priorities for detailed evaluation

44. The guidance provided in this note is premised on the need for revenue bodies to have a formal, structured and comprehensive approach for evaluating the impacts of their compliance risk treatments. However, revenue bodies need to be reasonably pragmatic in its practical application bearing in mind the limits of their resources and the relative importance/ significance of their individual risk treatment strategies. ***This applies both to the amount of effort expended up front in planning an evaluation as well as to the extent of effort and judgment expended in completing the final analysis phase of any evaluation conducted***. [11]

45. Revenue bodies have many compliance risks to be treated and operate with finite resources. This evaluation "challenge", which applies to all revenue bodies, is characterised in the following quotes.....

    "Developing an approach for evaluating program outcomes can become a very challenging task, one that involves more art than science. An appreciation of the technical strengths and weaknesses of various possible strategies for gathering evidence must be combined with an appreciation of the environment within which the evaluation takes place. This balancing must be done within the constraints imposed by limited resources and time. A combination of research and management experience is clearly required" (Treasury Board of Canada, Program evaluation methods, 1998)

    "Accurate impact measurements are costly in terms of resources that will in the end be diverted from the activities subject to measurements. They also require core staff with very specific skills that are in short supply in most (if not all) revenue bodies. This need not be a problem if only a few major impact measurements are carried out in areas subject to special attention. But a far greater degree of pragmatism is required if impact measurements are to be carried out on a broader scale, as we have opted for in the Danish compliance risk management framework. This is why we have adopted a pragmatic approach with extensive use of indicators and focus on establishing an explicit and (relatively) solid case for the treatment design." (Danish revenue body comments for this note).

46. In addition to the abovementioned factors, other considerations come into play:

    - Full evaluations for all initiatives could very quickly start to interfere with each other or interfere with normal compliance activity [12];

    - Sometimes the nature of an initiative makes it very difficult to use the experimental design that would be required to attribute effects. This is particularly true when using communications / marketing where it is very difficult to constrain the effects to a treatment group. It can also be true where an intervention involves agents – who will tend to communicate the intervention beyond the treatment group;

    - It is often the case that a number of initiatives are being introduced that are intended to tackle the same problem but from different angles (particularly so when funded from a central change programme). It can be extremely difficult to disentangle the separate effects;

[11] The guidance in this note describes what is termed a 'full scale' evaluation. However, for reasons that are explained in this part it is not feasible to conduct such evaluations for all compliance risk treatment strategies. For such strategies, a less intensive approach (referred to as 'lighter touch' approach) should be adopted.

[12] UK HMRC made the observation here that some years ago its analysts worked out the implications of putting in place control groups to measure the full impact of all major risk projects. The sums suggested that HMRC would quite quickly generate the situation where the need to maintain control groups would seriously constrain normal compliance activity.

- An intervention is thought to be so potentially valuable that it is immediately rolled out to the whole population – so preventing evaluation of the full effects;

- A considerable amount of organisation and discipline is required to fully operate experimental designs – in particular the need to preserve control groups if used.

47. These important considerations suggest the need for care and a selective approach when deciding which treatment strategies will be subject to full scale evaluations and those that can be subject to a less intensive and resource-demanding effort, a so called '*lighter touch'*. Concerning this 'lighter touch' approach, the most difficult part of evaluation is often to prove the causality between the observed change in behaviour and the activities undertaken by a revenue body. A pragmatic (lighter touch) approach could mean that, for some strategies cases, it is good enough to simply measure a change in behaviour without establishing the reasons for it.

48. An additional consideration concerns the care that should be taken to avoid a situation where only 'easy to measure' activities are undertaken, described by one member country in the following terms…… "*we are concerned that the activities that lend themselves to fulfill ideal evaluation criteria described in the note, will not necessarily represent the most effective activities – they might be chosen just because they enable good indicators to be found and there might actually be a risk that the more focus you put on the process of finding good indicators and minimizing the attribution problem the more you will use activities that lend themselves easily to be measured at the expense of the more effective activities"*.

49. For the reasons outlined, it is essential that revenue bodies have an effective process for considering and deciding evaluation priorities, particularly for deciding which compliance strategies are to be subject to full scale evaluation and those that can be evaluated in a less intensive and costly way. Such a process should consider, on the one hand, the possible consequences of not conducting full scale evaluations for some risk treatment strategies and, on the other, the benefit/advantages of a '*lighter touch*' approach. To assist in this respect, the note sets out hereunder the sorts of criteria that might be applied, either individually or collectively, in reaching a decision on which strategies should be subject to full scale evaluation:

    - The compliance risk to be treated by the strategy has been assessed as significant in a tax revenue/ compliance context;

    - The compliance risk treatment strategy entails a relatively large commitment of revenue body resources;

    - The compliance risk being targeted has a high media profile (e.g. concealing of income in offshore jurisdictions) and there is an expectation that the revenue body will make known publicly the impacts/ results of its work;

    - The risk treatment strategy involves one or more novel or unusual risk treatments that justify evaluation, to the extent practical, to inform future planning;

    - The compliance risk treatment includes changes to legislation (some times controversial) and there is a need to inform Government of its impact;

    - The compliance risk treatment is one that, if not addressed, could be the source of serious reputational risk for the revenue body, and even Government; and

    - Other factors/ considerations deemed important by the revenue body.

50. Suffice to say, decisions regarding which compliance strategies are to be subject to full scale evaluation (i.e. the priorities for evaluation) must take account of the availability of expert resources to support the work, should be made at the appropriate

organisational level (ideally, at a central level) and should be clearly communicated to all stakeholders.

The timing of evaluation action

51. Identifying the precise point in time when an evaluation should occur is based on when a revenue body expects to see an effect. In deciding when to undertake an evaluation, the following questions should be considered by a revenue body:

    - When does it expect to start seeing the change?
    - When does it expect the full effects of its strategies to be seen?
    - How often will it need to do an evaluation to be able to understand whether the change has been sustained?

52. The remaining part of this chapter describes the evaluation methodology —in essence, phases 3 and 4 of the methodology— in three parts (supported by reference to actual case studies provided by member revenue bodies):

    - Developing a suite of indicators for each compliance risk treatment strategy;
    - Evaluating the effectiveness of the strategy; and
    - Reporting on the effectiveness of the strategy.

Developing a suite of indicators for compliance strategies

53. Before a revenue body can determine whether its strategies have had the desired effect in treating a specific compliance risk, it must identify potential indicators, validate those indicators, determine a suite of indicators that will provide a credible picture of the impacts achieved, and decide how it will measure its chosen indicators. These activities should be taken soon after the risk statement, desired outcomes and strategies have been endorsed by the revenue body's internal risk management processes. Working through these steps will allow a revenue body to put the processes and tools in place to ensure that the relevant data can be collected and analysed.

54. There is usually a considerable time lag between setting up and actually doing an evaluation because a revenue body's strategies will take time to have an effect. However, if a revenue body is evaluating effectiveness retrospectively it will be in a position to undertake these activities almost immediately after completing phases 1 and 2 of the methodology.

Indicators

55. An indicator is not a measure – it is a signpost to help staff understand whether the strategies adopted to treat a specific compliance risk have resulted in the change(s) being sought. An indicator typically forms part of a suite of indicators chosen to help develop a credible picture of a strategy's overall effectiveness.

56. When developing a suite of indicators for each risk treatment, it is not essential to develop different indicators for each success goal. There may, in fact, be one or more indicators that can be used against several success goals. This can be helpful in reducing the overall cost of tracking progress.

Identifying potential indicators

57. As a starting point, a list of potential indicators should be identified that will show to what extent a revenue body has achieved its success goals.

Who should be involved?

58. This step should involve a range of subject matter experts to explore all possibilities including staff who understand the risk and the desired outcomes. They should also be aware of the success goals and strategies identified in phases 1 and 2.

59. Staff who have the relevant business knowledge and those familiar with the business data should be brought together, including the risk manager, intelligence experts, strategy experts, risk analysts, and business and data analysts.

Identifying indicators

60. Appropriate indicators should be identified by looking at each success goal and thinking about the changes one would expect to see if the strategies were effective and how they could be measured. The following examples may help consideration of this:

    - How would one know if there was an improvement in voluntary compliance?
    - What would tell a revenue body if there was a change in community confidence?
    - How did the revenue body know there was a problem in the first place?

61. When identifying indicators consideration should be given to whether they are capable of showing:

    - ***Sustainability:*** Identify indicators that can demonstrate progress and whether a positive change is maintained or declines over time. Select indicators that can show change over the immediate, intermediate and long terms.

    - ***Ripple effects:*** Include indicators that can show whether the strategies have had an effect on the wider population beyond the part of the population directly targeted by the strategies.

    - ***Unintended consequences:*** Consider indicators that will show whether any unintended consequences have emerged as a result of the strategies used.

62. As a general rule, a combination of indicators should be identified that focus on both the broad and the detailed levels to help identify whether there has been an observable change in compliance behaviour and community confidence. (NB: Case study examples 1 and 2 are particularly good examples of risk areas where a broad mix of indicators has been developed, admittedly involving some difficult-to-measure aspects.) This can also help explain whether a revenue body's compliance strategies are responsible for the change. It is advisable to avoid choosing only indicators that are easy to measure—even though an indicator can't be measured at this point, it may be useful at a later date when more data becomes available. In addition, discussion between staff from diverse backgrounds and capabilities can be invaluable in identifying indicators that might otherwise be overlooked [13].

[13] UK HMRC observed that repeatedly it has found it important to have indicators in place for each stage of the process of an initiative. For example, if a new piece of guidance is introduced it is important to measure awareness of the product as well as indicators of reducing error. This may sound obvious but examples have been observed of initiatives which did not produce an impact on the final outcome indicators and where it was then not possible to go back and diagnose why the anticipated success did not occur. Process indicators can also provide more confidence that any observed effects in outcome indicators are plausibly associated with the initiative. It has also found that social research, often used for our larger campaigns, can help understand why the quantitative outcomes realised occur. Although it does not produce robust statistics it can be crucial for digging into key implementation issues and giving the research extra validity.

Characteristics of indicators

63. When developing indicators, it is important that they are:

- ***Measurable:*** Indicators should be based on a combination of quantitative and qualitative data. Quantitative data is information that can be counted. For example.... "*78% of individual taxpayers filed their tax return on time*" provides a direct/ quantified measure of the proportion of returns filed on time, a measure of filing compliance.

  Qualitative data is largely descriptive data that is difficult to measure in numeric terms (although there may be quantified element attaching to it). For example.... *61% of taxpayers with incorrect claims indicated that their claim was incorrect because they did not understand their obligations*". This is sometimes referred to as the prevalence of a non-quantified attribute. In the example given, the data points to *why* non-compliance occurred but does not quantify *how much* (its impact) in revenue terms.

  In other words, quantitative data can tell you *how much* of something has happened while qualitative data is often extremely useful in explaining *why* it happened. This may also help you learn more about each risk.

- ***Expressed in a neutral form:*** The purpose of an indicator is to identify change. It should make no assumptions about the direction the change should take. In other words, express indicators in terms of a 'change' rather than a 'reduction' or 'increase'. (For example, an indicator for a specific 'failure to correctly report' risk could be 'a <u>change</u> in the average tax rate of targeted taxpayers' rather than 'an <u>increase</u> in the average tax rate of targeted taxpayers'.)

64. Examples of actual indicators used by revenue bodies and reported in case studies described in Annex 1 of this note are set out in Table 1 below.

Table 1. Examples of potential indicators from case study examples

Risk type	Indicator	References
Failure to file timely returns	Change in the timeliness (e.g. average number of days filed late of returns filed by targeted group of taxpayers	Case studies 3 , 5, 12
	Change in the timeliness of returns filed by targeted taxpayers vis-à-vis untargeted taxpayers	Case study 7
	Change in proportion of returns filed on time (and/or some other set date) by targeted group of taxpayers	Case study 4
Failure to correctly report liabilities	Change in level of liabilities reported in returns by targeted industry vis-à-vis other (untargeted) industries	Case study 1, 7
	Change in reported net income or average rate of tax of targeted taxpayers vis-à-vis other (untargeted) taxpayers	Case studies 7 , 8, 9
	Change over time in average tax rate or reported profit ratios of targeted group of taxpayers	Case studies 2 , 5, 11, 12, and 15
	Change in numbers of taxpayers reporting a specific item of tax return information (e.g. property ownership)	Case studies 6 and 9
	Change in amount of taxes withheld at source by third parties and remitted by taxpayers in targeted industries	Case studies 3 and 7
	Trend in the nature of queries/ ruling requests re a specific tax issue or arrangement subject to treatment	Case study 2
	Number of voluntary disclosures from a treatment strategy	Case study 6
Failure to pay taxes on time	Change in timeliness of taxes paid by targeted taxpayer group	Case study 3
	Change in timeliness of taxes paid by targeted taxpayer group vis-à-vis other taxpayers	Case study 16
All risk types	Change in taxpayers' perceptions of the revenue body's	Case study 11

	ability to deal effectively with the cash/black economy /1	
	Change in taxpayers' expressed degree of confidence in the competence of the revenue body /1	Case study 14
	Change over time in number of reports received from citizens concerning alleged non-compliance /1	Case study 1
	Trend in the nature of feedback from representatives of tax intermediaries generally or for a specific issue /1	Case study 2

/1. Measured by broad or targeted surveys of taxpayer and/or tax intermediaries' populations or sub-sets thereof.

Aligning indicators to success goals

65. The purpose of evaluation is to understand the extent to which the success goals and ultimately the desired outcomes have been achieved. Consequently, the indicators chosen must be directly linked to what a revenue body is trying to achieve.

66. Table 2 provides examples of aligning indicators to success goals. It also shows how the indicators can work together to help understand the extent to which the success goals have been achieved. Further examples can be found in the case studies at Annex 1.

Table 2: Example—aligning indicators to success goals

Success goals	Potential indicators
The community is less tolerant of non-compliant behaviour	• Change in the number of reports received from citizens etc containing allegations of tax non-compliance • Change in the responses to questions relating to non-compliant behaviour in a revenue body's community, business and/or tax practitioner perceptions surveys • Change in the tone of media comment re non-compliant behaviour
Taxpayers are more compliant with their return filing obligations	• Change in the percentage of the citizen or business population that file their tax returns on time and/or by a set period after relevant return filing due dates

Documenting the process

67. There should be documentation of the consideration given to each indicator including: 1) what the indicator is supposed to do; 2) the expected result if the strategies are successful; and 3) the external factors that may affect the usefulness of the indicator. This will help to ensure that what the indicator actually measures is in line with the original intent. It also ensures that these considerations are available to others who may work with the indicators at a later date.

Validating indicators

68. Once a potential suite of indicators for a compliance strategy has been identified each indicator needs to be critically examined to understand whether it will provide the information that is expected. This process is called validation.

Validation tests

69. Indicators can be validated by applying a range of tests, the main ones being:

- ***Attribution:*** Indicators need to measure something that the strategies employed can reasonably be expected to influence. Consideration should be given to whether a change in behaviour is likely to be the result of the strategies or whether changes may occur as a result of other factors. It should be borne in mind that: 1) behavioural change may occur regardless of the compliance strategies employed; 2) the strategies employed may only partially influence subsequent change, given other influencing factors; and/or 3) there may be no obvious change in behaviour because the strategies used only halt deterioration or maintain the status quo.

- ***Comparability*:** Continuity is necessary in order to make comparisons in trend analysis over time. While it is not always possible to avoid differences in the underlying data, recognition of those differences and the influence they have on the indicator analysis should be documented.

- ***Able to be substantiated:*** The results and conclusions drawn during the analysis of an indicator should be capable of being substantiated by an independent authority. In other words, an independent reviewer with relevant qualifications should be able to come to the same conclusions

- ***Free of bias:*** The information used to help measure effectiveness should be capable of being impartially collected, analysed and evaluated.

70. In addition, there are tests in the so-called 'SMART test' grouping [14]. This group of tests explores whether the indicators are specific (S), measurable (M), achievable (A), relevant (R) and timed (T). They are elaborated at Annex 5.

Attribution

71. As mentioned elsewhere in this note and highlighted by the guidance reflected in the material in Box 1 (from the Treasury Board of Canada's program evaluation guidance materials), attribution will be a frequent challenge, and this applies particularly in a tax compliance context. As emphasised by the Board in its guidance…….. "*Making causal inferences about results in evaluation means rejecting or accounting for rival possible explanations…………. One does this by setting up good comparisons that avoid as many threats to validity as possible*". Guidance on setting up good comparisons is set out at para. 77 et seq.

Box 1. Guidance on the issue of attribution

Often, a program is only one of many influences on an outcome. In fact, deciding how much of the outcome is truly attributable to the program, rather than to other influences, may be the most challenging task in the evaluation study.

The key to attribution is a good comparison. In laboratory settings, rigorously controlled comparison groups meet this need. In the case of federal government programs, less rigorous comparisons are generally possible and may be subject to many threats to *internal validity* and to *external validity*.

The following are the most common such **threats to internal validity**:

- **History**-events outside the program that affect those involved in the program differently than those in comparison groups;
- **Maturation**-changes in results that are a consequence of time rather than of the program (such as participant aging in one group compared with another group at a different stage);
- **Mortality** - respondents dropping out of the program (this might undermine the comparability of the experimental and control groups);
- **Selection bias**-the experimental and control groups are initially unequal in their propensity to respond to the program;
- **Regression artefacts—**pseudo-changes in outcomes occurring when people have been selected for the program on the basis of their extreme scores (any "extreme" group will tend to regress towards the mean over time, whether it has benefited from the program or not);

[14] **SMART** is a mnemonic for a practice used in project management at the project objective setting stage. It is a way of evaluating the objectives or goals for an individual project. The term is also in common usage in performance management, whereby goals and targets set for employees must fulfill the criteria. The first known uses of the term reportedly occurred in the November 1981 issue of Management Review by George Doran, Arthur Miller, and James Cunningham (ex www.wikipedia.org/wiki/SMART_(project_management)

- **Diffusion or imitation of treatment**-respondents in one group become aware of the information intended for the other group;
- **Testing**-differences observed between the experimental and control groups may be due to greater familiarity with the measuring instrument in the treatment group; and
- **Instrumentation**-the measuring instrument may change between groups (as when different interviewers are used).

Several **threats to external validity** also exist, which means that there are limits to the appropriateness of generalizing the evaluation findings to other settings, times and programs. In the federal government context, external validity is always a major concern since evaluation findings are usually meant to inform future decision-making.

Three groups of threats to the ability to generalize findings exist:

- **Selection and program interaction**-effects on the program participants are not representative because of some characteristic of the people involved (that is important to effects) is not typical of the wider population;
- **Setting and program interaction**-the setting of the experimental or pilot program is unrepresentative of what would be encountered if the full program was implemented; and
- **History and program interaction**-the conditions under which the program took place are not representative of future conditions.

It is obviously very useful in selecting evaluation strategies to be aware of the likely threats to validity. Much of the ingenuity in evaluation design, and in the ensuing data collection and analysis, lies in devising ways of establishing the effects attributable to the program. One does this by setting up good comparisons that avoid as many threats to validity as possible.

Source: Treasury Board of Canada, Program Evaluation Methods, 1998

Understanding strength

72. Indicators are not used in isolation to understand the extent to which a success goal has been achieved. When considered together as part of a suite, they should provide a credible picture of the outcomes.

73. Weak indicators sometimes perform a useful function in the overall suite and should not be discounted on their strength alone. However, the use of too many weak indicators —see example in Box 2—may make it difficult to present a credible picture of impacts that have occurred as a result of the strategy. This can be addressed by exploring ways to strengthen or replace some of the indicators.

Box 2. Examples of weak and strong indicators

Success goal: The community displays a reduced tolerance for participation in the cash economy.

Weak indicator: A change in the frequency and tone of media comment relating to the cash economy. This indicator is weak because it is likely to be too broad to identify any real behavioural change. Measuring something more specific can often strengthen the indicator.

Strong indicator: A change in responses to specific questions in a revenue body's surveys (including community, business and tax practitioner surveys).

Accepting or rejecting indicators

74. The validation process enables informed decisions to be made as to whether an indicator should be:

- *Accepted:* Where it is decided to accept an indicator that has validation issues careful consideration should be given to the impact these flaws will have on the strength of the indicator, as well as on its value in the overall suite of indicators.
- *Rejected:* Where it is decided to reject an indicator, consideration should be given to the impact this will have on the usefulness of the overall suite of indicators. If its absence is likely to have a detrimental effect, other alternatives should be considered.

- *Retained for future use:* Generally, indicators that don't meet all of the validation requirements should be discarded. However, if the flaw is only temporary, such as where a time-based indicator hasn't been in existence long enough to be able to track progress, then it should be retained for consideration in any future evaluation.

Painting a credible picture of a strategy's impact/effectiveness

75. Once the validation process has been completed, an assessment needs to be made to determine whether the proposed suite of indicators is likely to provide a credible picture of a strategy's impact/effectiveness. This should include consideration of:

    - Is the suite of indicators, in an overall sense, likely to provide sufficient evidence to justify conclusions regarding the extent of effectiveness of the risk treatment and/or a picture that is strong enough to withstand scrutiny and challenge?

    - Will the indicators show that there has been a change in behaviour of the risk population, there has been a change in community confidence, the change has been sustained, the strategies employed have contributed to the change, the effect can be seen more broadly than those targeted by the strategies employed, and were there any unintended consequences as a result of the strategies?

76. If the answer is 'no' to either question, the process of indicator identification should be extended to identify more potential indicators, and then to validate them.

77. Once the suite of indicators has been finalised, endorsement should be sought from the relevant decision making group in a revenue body.

Measuring indicators

78. Once it has been decided which indicators are to be used, it needs to be determined how they will be measured. For example, what method will be used to identify whether: 1) a change in behaviour or community confidence has occurred; and 2) the change can be attributed to a revenue body's strategies?

Identifying a change

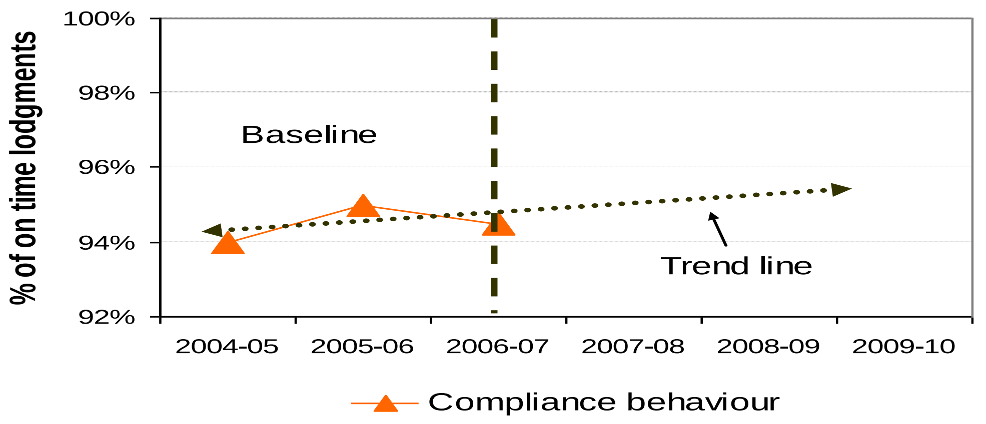
79. Before one can understand whether there has been a change in the behaviour or community confidence, it is necessary to know what the compliance behaviour or community confidence looked like before the strategies were implemented.

— Baseline

80. A baseline is a reference point against which other things can be compared or measured. It is used to identify whether there has been a change over time. (Baselines are also useful in establishing meaningful targets or measurable objectives.) A baseline measurement should reflect the starting point prior to the strategies being implemented. It should then be used for comparison purposes to determine whether there has been a change over time. Comparing against a single data point, however, is not always a reliable method of identifying a change. While there may be a difference between an original measurement and a later observation, there is no way to understand whether this change is caused by a revenue body's strategies or whether it simply represents natural fluctuations in the data being observed. Observing historical trends in the data before strategies are implemented will help identify whether a change has occurred.

81. Figure 6 looks at on-time annual income tax return filing behaviour for three successive fiscal years. It is important to note the general direction evident in the data (which is based on a three year average of actual filing rates).

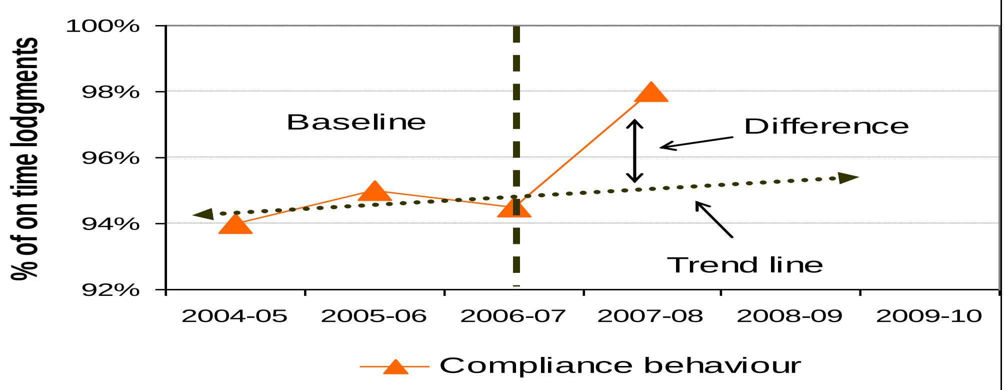
FIGURE 6: Understanding historical behaviour



—Identifying trends

82. A trend is the general direction in which something tends to move. Trend analysis involves the collection and presentation of information in order to identify patterns. Trend data are usually presented as a set of data points measured over uniformly-spaced, successive time periods. A trend will be evident where the data show a general tendency to move in the same direction over time. By drawing a line through the historical data to highlight the trend and extending that line over future observation points, it is possible to get a sense of where the subsequent observations should be if the trend is maintained and there is no change. The example in Figure 7 shows a difference between the trend line and the new data point. This indicates that something has affected the behaviour.

FIGURE 7: Using a trend line to understand normal patterns



83. To understand historical trends, the more observations that are available, the easier it will be to identify patterns. A longer time frame can often help to identify less obvious patterns such as those presented in seasonal data. In a taxation context, collecting reliable data over a longer period is problematic because it introduces a range of other variables that can affect the observations particularly when dealing with annual tax data. These variables include changes in the way the data are collected, differences in the timing of the intervals between data collection, changes in the law, changes in the way the law is being administered, economic turbulence, and societal changes. While it is preferable to have enough observations to make the claim of a trend valid and reliable, having less does not rule out the indicator as a useful contributor to the overall suite of indicators. It can still add valuable insight and support for other indicators.

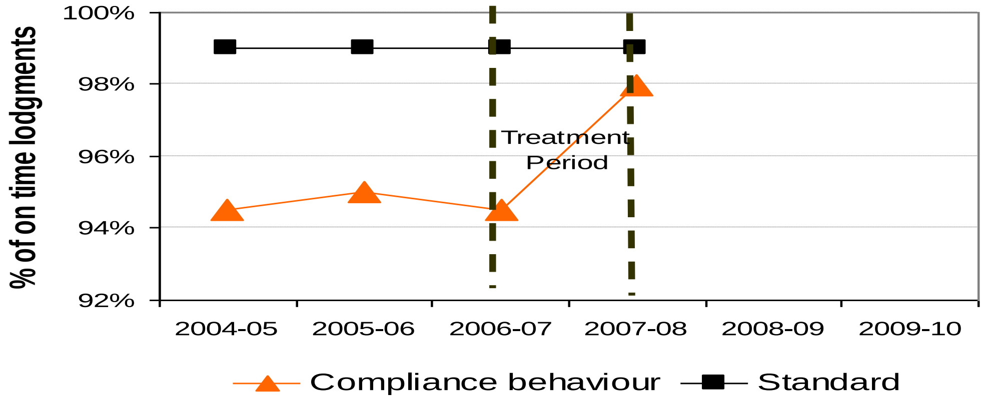
—Analysing trends

84. Several methods can be used to understand whether there are changes in trends:

- *Pre- and post- studies:* Pre- and post- studies assess changes as a result of an intervention. Measures are taken both before and after the intervention in an effort to understand whether there has been a change. For example, tracking behaviour over time, both before and after the compliance strategies are implemented, can help staff to understand whether there has been a change in behaviour that coincides with the implementation of the strategies. But it will not indicate whether the strategies caused that change, although in particular circumstances such changes can reasonably be inferred (with supportive information)—see Table 1 for case study references.
- *Longitudinal studies:* A longitudinal study involves repeated observations of a specific set of participants over a period of time. It can allow a revenue body to measure key variables at different points in time. The same data are collected from the same participants at regular intervals. Longitudinal studies are particularly useful for understanding the long-term effects of a revenue body's strategies.

  When choosing a sample for a longitudinal study, consideration should be given to the likely effect of 'churn' on the sample size. Churn is the number of people who move in and out of a particular group over any given period of time. If a sample population has a high churn rate it is likely a significant number of participants will drop out along the way. Increasing a sample size to allow for natural attrition of the participants over time is one way to overcome this.
- *Measuring against standards/targets:* A standard or target is a reference point against which compliance behaviour or community confidence can be measured. It reflects a required or desired level of performance that indicates quality or success. For example, measuring behaviour against a standard or target allows a revenue body to see whether behaviour is moving towards (or has realised) what it regards as an acceptable standard of performance (see Figure 8).

FIGURE 8: Comparing against a standard or target



A number of revenue bodies in OECD member countries use standards/ targets in a tax compliance context—see Annex 2 for examples.

Attribution

85. Identifying a change in compliance behaviour or community confidence does not necessarily mean that a revenue body's compliance strategies caused that change. Using a comparison technique enables it to separate the effects of its strategies from other factors that could be causing the change. This will help it to identify whether the

observed change is a result of something it has done or a result of other influencing factors that are outside its control. Several different types of comparisons can be used. Some are suitable for high level measurements (for example, benchmarks) while others are more suitable for drilling down to a sub-population level (for example, target and control group studies).

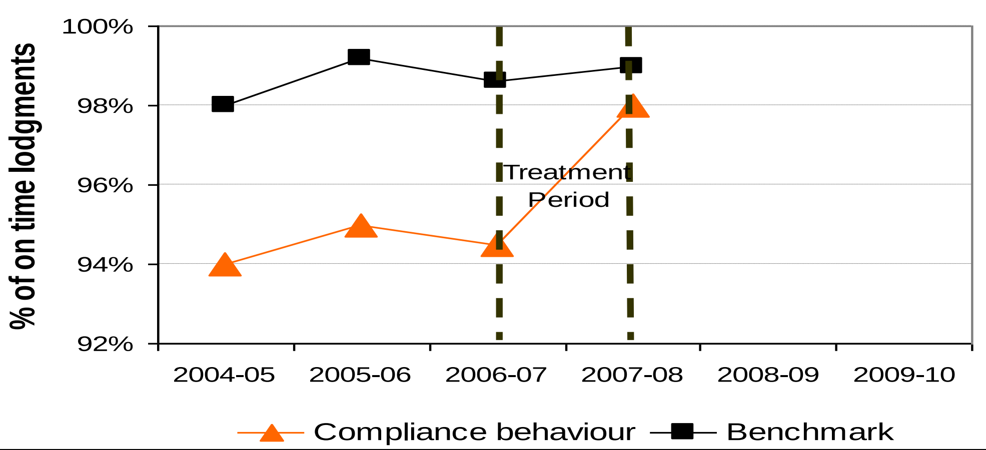
Benchmarks

86. A benchmark is a reference point for comparison purposes. It isolates the effects of the strategies from the effects of external influences such as economic fluctuations. For example, comparing the behaviour of a risk population to an appropriate benchmark is useful in understanding whether the observed behaviour is in line with expectations. If a revenue body's strategies have had an effect, it would expect to see a change in the relationship between the benchmark and the compliance behaviour of the risk population.

87. Benchmarks are often based on external data, such as statistics produced by central statistical agencies, and are comparable across the broader population base. They are particularly useful when examining behavioural change at the broader level, but may become less useful when drilling down to a specific sub-population.

88. Figure 9 shows that both the benchmark and the population of interest were tracking together until the start of the treatment period. The measure taken at the end of the treatment period shows a distinct difference in the behaviour pattern of the two groups.

FIGURE 9: Comparing against a benchmark



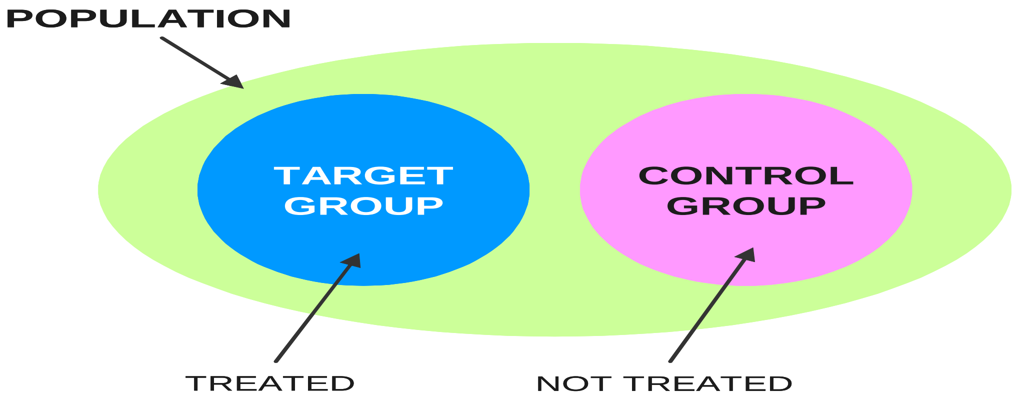
Inferential statistics

89. Inferential statistics are a group of methodologies that allows one to draw inferences about a population based on the behaviour of a sample. When an appropriate benchmark is not available, inferential statistics can be used to identify whether or not there is a causal relationship between the behaviour being observed and the strategies being used. They use sampling methodologies to identify target and control groups from within a specified population (see Figure 10).

- *Population* – A group of taxpayers who share a common set of characteristics.
- *Target group* – A sub-group of taxpayers selected from the population and used as the focal point for the strategies.
- *Control group* – A sub-group of taxpayers selected from the population that is isolated from a revenue body's strategies for the purposes of the study.

90. Figure 10 illustrates the relationship between the population, the target group and the control group.

FIGURE 10: Target and control groups within a specified population



91. Comparing target and control groups from the same population allows one to screen out the effects of unknown variables. Theoretically both groups will be subject to the same environmental factors except for the compliance strategies. Any differences in behaviour can be attributed to the strategies being used.

92. Methods for comparing target and control groups include randomised control trials and the use of matched groups. These are described briefly hereunder and more fully in the accompanying compendium note:

- ***Randomised control trials*****:** Both the target and the control groups are randomly selected from within the population that is to be targeted by the strategies. The control group is then isolated from treatment. As the target group has been subjected to the treatment and the control group has not, any observed behavioural differences between the two groups can be attributed to the compliance strategies.
- ***Matched groups*****:** When it is not possible to use a randomised control trial, a matched group can be used instead. This is a comparison group that is similar to the treatment group. Groups are matched according to similarities in their characteristics. While the results from a matched study are adversely affected by unknown differences between the groups, well-matched groups can approach the rigor of a randomised control trial.

93. Use of the abovementioned (and other) methods by revenue bodies in OECD member countries has grown in recent years in line with the adoption of improved compliance risk management processes, and is evident from a number of case studies provided with this guidance note—see Table 3 below and detailed case study examples in Annex 1.

Table 3. Case study examples using special methodologies

Country case study	Target industries/areas	Compliance risks /1	Methodology
Australia	Filing of high tax level income tax returns	2	Longitudinal study
Canada	Construction	2, 3 and 4	Randomised control trials
	Trucking	2, 3 and 4	Randomised control trials
	Restaurants (employee tips)	2, 3	Pre- and post-study
Denmark	Owners of shares & related profits on share sales	3	Randomised control trials
	Owners of foreign real estate and	3	Pre- and post- studies

	driving school instructors		
Singapore	New GST registrants	3	Longitudinal study
	Real estate agents	3	Pre- and post study, longitudinal study
Sweden	Payrolls of hair-dressers & restaurants	3	Pre- and post- studies
United Kingdom	Persistent VAT debtors	4	Randomised control trials
	High income earners—social contributions & PAYE withholding avoidance	3	Pre-and post- studies

/1. 1- Registration; 2- Timely return filing; 3- Correct reporting; and 4- Timely payment

Seeking expert assistance

94. Assistance in understanding the broader environment or identifying appropriate external benchmarks can often be obtained through internal specialist areas of a revenue body (e.g. research units, revenue forecasting and monitoring functions) or economic specialists in operational areas. Ideally, expert assistance will be engaged as early as possible in planning for an evaluation—it is often not possible to do proper evaluation on a retrospective basis.

Identifying necessary data requirements

Who's responsible?

95. Action should be taken to allocate ownership of the data gathering and indicator analysis tasks. Making such expectations clear from the outset should allow a revenue body to identify and deal with potential issues that may impact the timely availability of the information.

What data sources will be used?

96. It is important to be as specific as possible so that staff will know exactly what data are required. A useful first step is to bring subject matter experts on the research and data analysis side together with subject matter experts on the business side so that the former can become familiar with the intended target. By doing their "homework" up front, researchers should be in a better position to select the correct data set to describe the characteristics of the target and this knowledge will subsequently assist in the selection of compliance indicators and the development of the data requirements. Issues to be considered when identifying data sources include:

    - Is the data from an internal or external source? Internal data are obtained from sources such as a revenue body's computer databases of taxpayer information and regular surveys carried out of taxpayers' perceptions. External data are obtained from sources outside of the revenue body, such as the central statistical agencies.
    - Are there any issues around ownership of the data that will affect the ability to acquire it? For example, data from property databases of other government agencies may require special permission before it can be accessed. Sufficient lead time should be allowed for that permission to be obtained.
    - Are there any issues with regard to the availability of the data that would affect its usability? For example, there is sometimes a significant lag time between the collection and subsequent availability of government population census data that restrict its usefulness as a benchmark.
    - When and how often do the data need to be collected? It is likely that data will need to be obtained at regular intervals to track progress over time.

- Can the data be replicated? The data may need to be replicated for various reasons including the need to provide the same information at specific intervals for time-based analysis, and for quality assurance processes.

97. There is often a cost associated with the collection of data. Potential costs to a revenue body can range from payments to an external organisation to the opportunity cost of diverting internal resources to the task of collecting the data. Potential costs to the community include an additional burden on the time and resources of respondents. Before making a final decision on the data source, the following should be considered:

- *Are there any existing data sources that will suit requirements?*

  Collecting new data is generally more expensive and time consuming than collecting existing data. There may be existing data sources such as previous evaluation reports, other agency data or other internal and external data, that will suit requirements. There may be a range of existing surveys and studies that are potentially suitable. However, in such situations it is important to understand the purpose of the original study and any limitations it might have in relation to a particular indicator.

- *Is the data in a form that is able to be analysed for a particular purpose?*

  Information which is collected at source (primary data) for the particular purpose of meeting the evaluation objective is more reliable than information derived from a pre-existing data source. This is because the primary data has been collected with the particular purpose in mind.

  Secondary data has usually been collected, summarised or interpreted for a different purpose and may not be entirely relevant when used as a substitute for primary data. This may weaken the reliability of the indicator being used. Before decisions are made to use a secondary data source, effort should be made to understand the purpose for which the data was originally collected, how the data was collected, the currency of the data, whether the data is representative of the population of interest, and whether the data can be used for evaluation.

Are there any data limitations?

98. Consideration needs to be given to identify any limitations or issues that will impact on the reliability of the underlying data. Limitations can include: 1) changes over time in the method of data collection or compilation; 2) unavailability of data for certain time periods; 3) possible time lags in the availability of data; and 4) the need to test the volatility in the data against the expected size of impact(s) [15].

What comparisons will be used?

99. It should be made clear which comparisons are to be used to make sense of the changes for each indicator. In making such decision consideration should be given to: 1) how the baseline will be measured; 2) if a benchmark is to be used, whether it will be based on internal or external information; 3) whether a control group and target group need to be selected; and 4) whether a similar group should be identified for comparison purposes. Once the comparison techniques to be used have been decided, action should be taken to identify what needs to be put in place to collect the data.

Documenting intentions

100. There should be a record made of information relating to the underlying data sources, data limitations and comparisons associated with each indicator. This provides a sound

[15] Concerning this issue, the UK HMRC observed that it had found that the volatility in returned profits for some types of businesses makes evaluation of leverage (using targeted communications) initiatives very difficult. This is an example of why expert assistance is required very early in the evaluation process.

basis for quality assurance purposes and allows for independent verification or replication at a later date, if necessary.

Evaluating effectiveness

Gathering data

101. Once sufficient time has elapsed for the strategies to have had an effect on voluntary compliance and/or community confidence, a revenue body needs to begin gathering the data to enable it to analyse each indicator.

Who's involved?

102. The staff/ area responsible for gathering data and analysing each indicator should have been identified during the indicator identification and validation processes. For some strategies, it may be necessary to have different staff members responsible for each indicator given the specialised knowledge required to understand and interpret data.

Data integrity

103. When collecting data, there should be a sound process in place to verify and validate it. This should entail consideration of the following sorts of factors:

- ***Relevance of the data:*** The data should actually measure the characteristic it claims to measure. This is particularly important if a data set is being used that has been created for another purpose. If using secondary data, there should be a rigorous review of data specifications to ensure that the data are appropriate for the purpose to be used.

- ***Data availability and collection***: Ideally, data should be collected in a form that is tailored to analysis requirements. This will not always be possible particularly where secondary data are used. If using secondary data, the underlying data properties should be explored to understand any limitations on its analysis.

- ***Consistency of data***: The data should be consistent over time. This allows for repeated observations to check whether the effects of the strategies are sustainable. If this is not possible, efforts should be made to look for ways to mitigate the limitations of the data in the longer term.

- ***Timing of data*:** Where data are sourced from external databases or reflect external market trends, steps should be taken to ensure there are no issues with differing time frames when making comparisons. If there are any differences, these must be accounted for in the time lag in the analysis.

- ***Expertise of data collectors:*** Data quality can be impacted by the skills and expertise of the people involved in gathering it. Expert advice may be needed to ensure the data collection techniques are appropriate.

Quality assurance

104. Data collection techniques must meet appropriate quality assurance standards. This should include: 1) any limitations of the data source (for example, is the survey poorly designed? are there sampling errors and have these been articulated?) 2) any assumptions made; and 3) the approach taken and whether it is appropriate (for example, is the data coding approach correct?).

Data storage

105. Once collected, data needs to be appropriately stored. Storage requirements vary, depending on the type of data. The key elements to consider are security of the data, particularly where it is classified by a revenue body as of a confidential nature, and ease of data retrieval. The data specifications should also be stored in a way that will allow the data to be reproduced at appropriate intervals for cyclical evaluations, ongoing trend analysis and quality assurance purposes.

Analysing the indicators

106. Once the data for each indicator has been collected, it should be analysed to understand where there is any evidence of a change in behaviour and/or community confidence. The results for each indicator should then be brought together to determine whether, as a whole, they provide credible evidence of the success of the strategies in meeting stated success goals and, ultimately, the desired outcomes.

Who's involved?

107. Indicator analysis should be done by business or data analysts who understand both the business and the underlying data. In many cases the analysts will need to work closely with other subject matter experts in order to provide a comprehensive analysis. These experts may include risk and strategy experts, economists and researchers.

Making sense of the data

108. To evaluate the effectiveness of the strategies employed, there must first be an exploration of whether there has been a change in compliance behaviour and secondly, whether a revenue body's strategies caused that change.

—Is there a change?

109. An easy way to identify a change is to use a line graph to explore patterns and changes. To do this, the following steps should be taken:

1) Plot the compliance data on a line graph, preferably using one that allows the plotting of more than one data series. This will make it easier to compare data with a benchmark, for example.
2) Identify the point in time when the strategies were implemented.
3) Look at the historic observations before the strategies were implemented and identify any obvious trends.
4) Draw a line that depicts that trend and extend this line into the future.
5) Examine any observations taken after the strategies were implemented and compare them with the previous trend. If there is a difference, it can be concluded that there has been a change.

110. Where the historical data shows considerable fluctuation it may be difficult in the short term to identify whether a change has occurred. There may be a need to continue monitoring over time to understand whether there is a difference.

—Did the strategies cause the change?

111. Once a change has been identified it must be determined, to the extent practicable, if that change can be attributed, to the strategies used or whether it is a reflection of other changes in the environment. Comparing the data gathered to something else, such as a benchmark or control group, will help to isolate the effects of the strategies from the general fluctuations caused by other factors in the environment. To do this, examine the comparison method identified earlier, plot the comparison data on the graph, and

compare the pattern of movements in both the comparison data and the compliance data to see if there are any differences. If the data show a difference in the pattern of movement after the strategies were implemented, it can reasonably be concluded that the strategies had an effect on compliance behaviour or community confidence. If, however, the compliance data display the same pattern as the comparison statistic, it can be concluded that the observed changes in behaviour were caused by environmental factors.

112. Comparison data can be generated through the use of benchmarks and inferential statistical techniques such as randomised control trials or matched groups, as discussed earlier in this note.

113. Finally, it is important to bear in mind that attribution should not be viewed as an absolute concept, aimed at acquiring irrefutable evidence of the outcome being sought. Generally speaking, for those strategies where comparison data may not be available or are incomplete, it will generally only be necessary to assemble sufficient evidence to support a case that can be considered as reasonable, credible, and/or plausible in the circumstances.

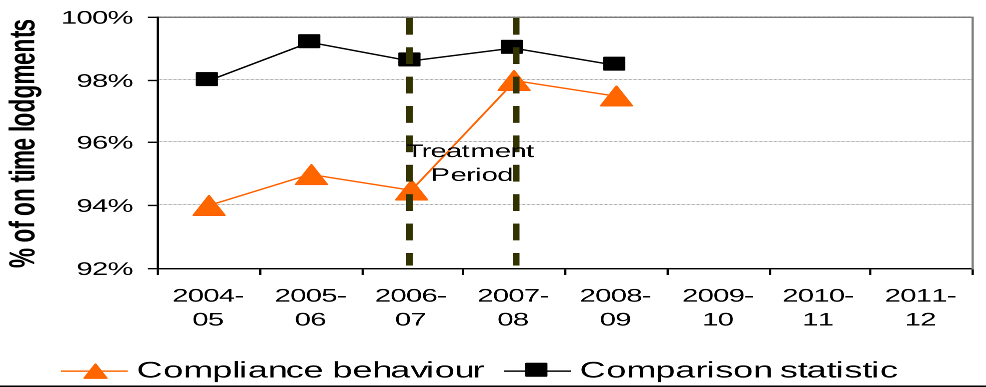
— Is the change in line with expectations?

114. Sometimes a revenue body's strategies can have a different effect than originally anticipated. Where there is an unexpected change, further inquiry should be undertaken. If potential unintended consequences have been anticipated from the outset, there may be indicators specifically designed to see whether these consequences have emerged. The emergence of unintended consequence(s) should not necessarily be viewed in a negative light. Some may, in fact, be positive and, even if negative, the insights they provide are valuable in reshaping compliance improvement strategies.

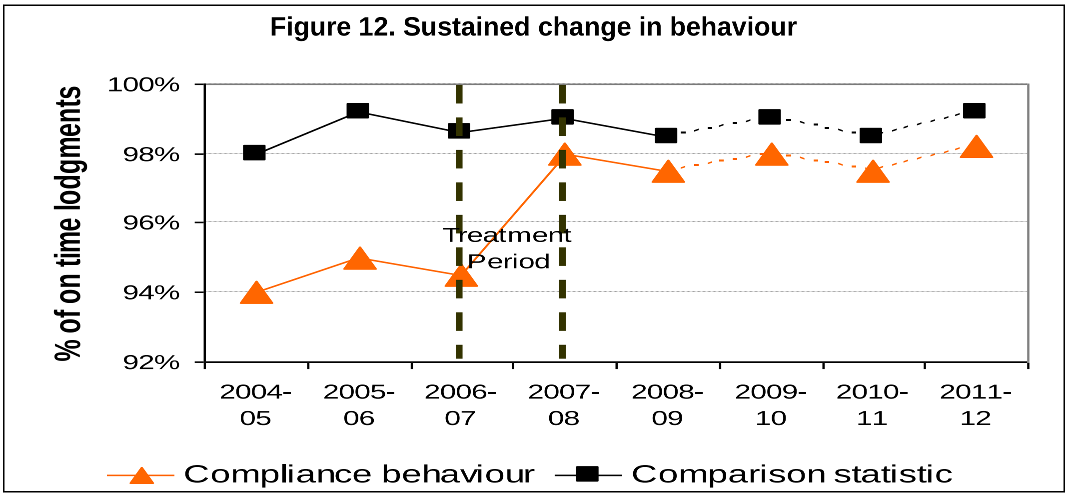
Identifying sustained change

115. A sustained change is one that lasts for the longer term. Identifying whether a behavioural change has been sustained needs to be done with repeated measurements over time. Figure 11 shows a change in the relationship between the compliance behaviour and the comparison statistic that corresponds with the treatment period. The 2008-09 observation indicates that the new relationship is holding.

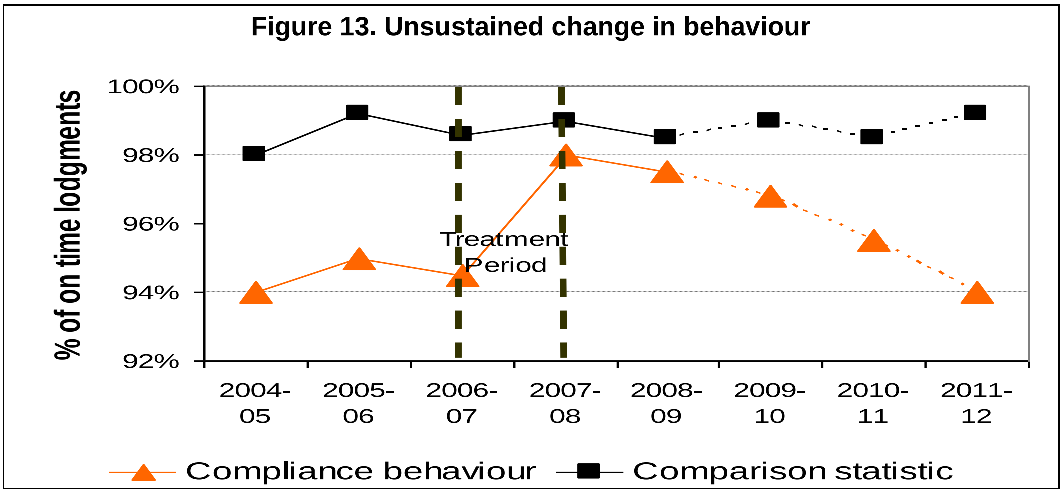
Figure 11. Identifying sustainable change in behaviour



116. If the behavioural change is sustained, there should be a path of the two groups continuing to track in parallel over an extended period of time, as seen in Figure 12.



117. Tracking the behaviour for a short period after the intervention will not always give a reliable indication of whether the effect is sustainable over time. As Figure 13 illustrates, while the behaviour is tracking in parallel with the comparison statistic in the 2008-09 income year, the slight decline in compliance behaviour continued. The declining behaviour was originally masked by changes in the environment. Looking at the pattern of observations over a longer period clearly shows that the effect of the strategies was short-lived and not sustained.



118. There is no set period of time over which the data should be tracked to identify whether a change is sustainable. However, a revenue body will typically need to track behaviour for several fiscal years to really understand whether its strategies have had a lasting, sustained impact.

Determining effectiveness

119. While the analysis of each indicator provides some useful insights, it will not provide a credible picture of the extent to which the strategies have improved voluntary compliance and/or community confidence. To build such a picture, **all** the analysis results for **all** of the indicators should be brought together and examined as a whole to determine the extent of effectiveness and draw credible conclusions. This is a critical phase of the evaluation process because often careful judgments will be required in order to properly synthesise the findings (including dealing with the issue of attribution), while also having to recognise that some of the indicators may be weaker than others and/or supported by data that is less reliable or complete.

Who should be involved?

120. Evaluating compliance effectiveness should be carried out by officials who understand both the risk and the environment affecting that risk. While the risk manager will generally take responsibility for making sense of an evaluation's findings, it is good practice to involve other subject matter experts to help make sense of the information as a whole.

121. As for the indicator analysis, the staff that should be involved in examining and interpreting the findings should include relevant risk and strategy experts as well as analysts who have an understanding of the environmental variables that can affect the risk area.

Bringing the case together

122. When the suite of indicators was originally chosen, a number of assumptions were made about what one would expect to see if the strategies were effective. Once the results have been brought together, these early assumptions should be revisited to see whether the indicator results are as expected.

123. If the behaviour is in line with earlier expectations, then determining the extent of effectiveness should be relatively easy. If any indicators show a different result, however, there should be an objective investigation of the results to understand how they impact on conclusions about the extent of effectiveness of the strategies used.

124. Where an indicator shows an unexpected result, the following issues should be explored:

- Is the result significant in relation to the overall suite of indicators? If the indicator is weak, its impact on the overall outcome may not be significant. If it carries a lot of weight within the suite, it needs to be understood how it affects the outcome.
- Do any other indicators show similar results? If the result is supported by a range of other indicators, then it would be reasonable to conclude that the strategies have not been effective.
- Is it an unintended consequence? Any unintended consequences must be explored in order to understand their impact on the risk and for planning future compliance strategies.
- Is further analysis necessary? Further analysis may be required to understand the reason for the result. Issues such as a flaw in the data that was not previously identified, or a defect in the original thinking, should be considered.

Drawing conclusions

125. Evaluating compliance effectiveness is based on understanding whether the compliance strategies have moved the revenue body towards achieving its success goals and ultimately desired outcomes. The evaluation literature [16] offers some guidance to the formulation of ***credible*** conclusions:

- To be objective and credible, the conclusions must follow the evidence; furthermore, credibility is, in part, a question of how the conclusions are reported—the believability of conclusions depends partly on how they are presented.

- Objectivity is paramount—this means that the evidence and conclusions can be verified and confirmed by people other than the evaluator; it should always be clear to the reader what the conclusions are based on, in terms of evidence

[16] Treasury Board of Canada, *'Program Evaluation Methods'*, 1998.

gathered and assumptions made; if conclusions are ambiguous, the underlying assumptions should be clearly spelt out.

- The conclusions reached should be based on a comprehensive coverage of the evidence /findings;

126. In practice, the indicators aligned to each success goal should be examined and decisions made as to whether, collectively, they show the extent to which a revenue body has achieved each goal. If the evaluation shows that the success goals have been met then it can be concluded with a reasonable degree of confidence that progress has been made towards achieving the desired outcomes.

Making recommendations

127. The lessons learned from evaluating effectiveness can help re-shape a revenue body's future compliance improvement strategies. In presenting recommendations, the following sorts of issues/ questions should be considered:

- Was the right mix of compliance strategies used to address the risk? If not, what mix might be better into the future?
- How could the design or execution of the strategies have been improved?
- Does more risk assessment work need to be done (for example, more research into a particular population's compliance behaviour)?
- Did any unintended consequences emerge? How might any negative outcomes be mitigated?
- How can the application of the compliance effectiveness methodology be improved (for example, better articulation of success goals, data collection methods or analysis techniques)?

Presenting a quality outcome

128. A quality evaluation of whether, and to what extent, the strategies have achieved success goals and ultimately the desired outcomes can provide a wealth of knowledge for a revenue body on which to base future risk treatment decisions, or even changes to tax legislation where such action might be deemed appropriate. To ensure a quality evaluation consideration should be given to the following sorts of issues:

- *The analysis methodology*—Choosing the right methodology is particularly important when looking at attribution issues. The methodology selected must be able to identify whether the changes observed are caused by a revenue body's strategies or by factors outside its control.
- *Any over-emphasis on one or more indicators*—While it is acceptable for some indicators to carry more weight in the overall picture than others, it is important not to overly rely on particular indicators simply because they tell the story that a revenue body wants to hear. The value from evaluating compliance effectiveness is not about success or failure but rather from the lessons learned and how they feed into the continuous improvement process.
- *Personal bias*—The temptation should be resisted to manipulate evaluation outcomes to paint only a positive picture, including by use of language that is not neutral in its tone (e.g. using language which attempts to overplay or downplay the significance of data). Evaluating effectiveness is not meant to be a statement of success or failure. It is meant to provide an understanding of whether a revenue body is on the right track with its strategies and to provide some guidance for future planning. Smoothing over any negative results can mask valuable lessons that could influence future risk treatment plans.
- *A credible case versus a definitive one*—Understanding what the indicators are really saying can be difficult due to the subjectivity inherent in arriving at a

conclusion. Evaluating effectiveness, particularly in a tax compliance context, is about presenting a credible case. It is not about providing a definitive measurement. Knowing where to draw the line between definitive and credible evidence is quite difficult because the level of comfort in drawing conclusions about effectiveness is intrinsically linked to an individual's personal opinion, knowledge, and experience etc. When deciding on the level of information needed to support the case being presented, thought should be given to how, if asked, the revenue body would justify its conclusions. Involving other experts with relevant knowledge in the evaluation can assist increase everyone's comfort levels.

Quality assurance

129. Involving the right people throughout the evaluation process can help to increase the quality of the case being presented. Depending on the significance of the compliance issue under review, it may be appropriate to have the claims independently verified to ensure that the conclusions are soundly based. This may be feasible in large revenue bodies where independent evaluation/ assurance functions are established and manned by staff skilled in such responsibilities.

Reporting on effectiveness

130. Evaluating compliance effectiveness presents an opportunity for continuous improvement based on an improved understanding of the risk. Presenting a concise report is an important means of communicating the outcome of the evaluation undertaken.[17] It is usually prepared by the designated risk manager in consultation with other subject matter experts.

When to report

131. The regularity of the reporting cycle can be affected by a range of issues including:

- *A revenue body's planning cycle* — Consideration needs to given to when a report will be needed for normal planning cycle purposes. Understanding standard revenue body planning and reporting cycles helps to determine when the report is likely to be required.
- *A revenue body's evaluation cycle* — The priority, pattern and severity of the risk may also require an evaluation to be undertaken on a more regular basis. Significant external factors such as volatility in the market place or the economic environment may also impact on the timing of an evaluation.
- *Other business reporting cycles*— Compliance effectiveness may be subject to other reporting requirements at the business unit and corporate levels (e.g. to meet a Government reporting need, to report to Government audit bodies).

132. The most appropriate time to produce and present the report should be decided after consideration of the abovementioned factors.

How to present the evaluation report

133. While the nature and size of a report will vary from case to case and jurisdiction to jurisdiction, it should be reasonably detailed for reference purposes, particularly if further action is to be taken for the specific risk area and/or new revenue body staff are to be involved. Also important is whether the report's findings might be used, for example, in the context of justifying a recommendation for legislative change, or to substantiate to an external audience the appropriateness of actions taken. There should also be a summary document prepared that presents the key points in a clear and concise form.

[17] In practice, a variety of media may be used to communicate findings depending on the strategy being evaluated and the intended audience.

134. The actual format used to present a report may vary according to the specific requirements of individual operational areas within a revenue body (e.g. large taxpayers, SME's, and debt collection). The content, however, should be broadly the same regardless of the intended audience or the required format.

The executive summary

135. The report should be presented in a way that makes the information easy to understand. The best way to do this is to present an executive summary supported by a more detailed report.

136. The executive summary (typically no more than two pages) should provide the reader with an understanding of your conclusions along with a brief outline of the evidence underlying those conclusions. It should also provide recommendations for the future. Specifically, attention should be given to the issues raised by the following questions:

- Has the strategy been effective? Include a brief summary outlining the conclusions about the extent to which the desired outcomes were achieved. This should include observations of whether there have been any ripple effects and whether the results have been sustained.
- What evidence has been provided to support your conclusions? The summary should provide an outline of the evidence that supports the conclusions. It should include brief details of any data limitations, assumptions and inferences that are relevant to the conclusions. It should also include a concise account of any contextual information, such as the effect that other factors may have had on the behaviour or the results, which will assist the reader make sense of the findings.
- What has been learnt? There should be a description of what has been learnt during the evaluation, including: 1) improvements in understanding the risk; 2) what treatments did /did not work as intended; and 3) unintended consequences.
- What further actions, if any, should be taken?

137. If appropriate, recommendations should be made for future risk treatment strategies, including such matters as the mix of strategies, the need for further risk assessment work, improvements to the evaluation methodology itself and the related process.

The detailed report

138. The summary should be accompanied by a more detailed report of the evidence that supports the conclusions. Specifically, it should include:

- the analysis results for each indicator, focussing on the specific behavioural shifts that occurred and whether these changes were evident in the wider population or restricted to the specific target group;
- a discussion of the evidence that supports the conclusion that the change in behaviour and/or community confidence can be attributed to the compliance strategies employed; and
- whether the suite of indicators demonstrates the extent to which each of the stated success goals and ultimately the desired outcomes have been achieved.

139. Where a positive shift in compliance behaviour or community confidence has been identified, the report should discuss whether:

- it is possible to identify a sustained change; it may be too soon to determine whether a behavioural change can be sustained over the longer term.
- any unintended consequences have emerged; suggestions for mitigating any negative outcomes in the future should be included.

140. Any recommendations made should include any findings concerning:

- which compliance strategies worked or did not work and any suggestions for improvements in the future;
- whether the right mix of strategies was used to address the risk;
- whether the risk is still occurring in the same way or whether it has changed; and
- whether more risk assessment work is necessary to better understand the risk.

141. Providing all of this information to a revenue body's decision-makers and planners will contribute to greater awareness of compliance related issues and inevitably improve decision-making concerning future compliance activities.

142. The final step entails the presentation of the final report. How this is done and who is involved will vary from one revenue body to another. For the more important and/or significant strategies some special form of reporting action should be considered. For example, in addition to submitting a written report, it may be desirable to have an oral presentation of the key results to senior decision-makers. This can also involve subject matter and research experts being present to elaborate where needed and to directly respond to questions about the methodology or the significance of the results.

V. Findings and Recommendations

143. As noted at the outset of this note, there is considerable interest in the development of practical approaches and methods for systematically evaluating the effectiveness of specific compliance risk treatment strategies. With many revenue bodies today under considerable pressure to improve tax compliance/ close the tax gap while operating with reduced resources, this interest now assumes even greater importance.

144. The guidance contained in this note draws on innovative work, based on considerable research, carried out by one OECD revenue body, and considerable further assistance from a number of other revenue bodies that have been intensifying their efforts to better understand the impacts of their compliance program activities. Work completed by the EC's Fiscalis Risk Management Platform Group has also provided valuable references. The methodology has been introduced to the broader FTA membership in presentations and discussions and there is agreement that it provides a sound and practical approach for broader application. Specifically, the note:

- describes a number of important concepts and issues that are often raised in an evaluation context (e.g. outcomes, attribution, and indirect effects);

- provides a description of a four phase compliance effectiveness methodology, expressed in high level terms;

- sets out step-by-step practical guidance for each element of the evaluation methodology: 1) developing a suite of indicators for evaluating compliance strategies; 2) evaluating effectiveness; and 3) reporting on effectiveness;

- emphasises the need for considerable pragmatism in deciding which risk treatment strategies should be subject detailed evaluation and, sets out as examples, the following situations where such action may be deemed appropriate:

  — the tax revenue at risk has been assessed as significant;
  — the intended risk treatment entails a large commitment of resources;
  — the risk being targeted has a high profile in the media;
  — the risk treatment strategy involves novel/ experimental actions;
  — the risk treatment includes legislative action; and
  — the risk, if not addressed, could be the source of serious reputational risk.

- provides a large number and variety of case studies from selected revenue bodies describing specific measures and indicators being used to evaluate the effectiveness of risk treatments and, for some, the impacts observed in practice.

- identifies in broad terms the benefits and costs associated with implementing and administering the recommended approach.

145. The guidance note is accompanied by two other documents that support use of the guidance—a background note '*Selected revenue body and other experience with implementing risk treatment evaluation*' and a companion guide '*Methodological techniques for use in evaluating the effectiveness*'.

Recommendations

- Revenue bodies are strongly encouraged to examine the evaluation guidance provided and to consider its relevance and practical application in their respective contexts; and

- Revenue bodies are strongly encouraged to have in place a sound and practical approach to evaluation that can be relied to provide credible evidence pointing to the effectiveness of their key compliance risk treatment strategies.

Annex 1: Case Study Examples

Case study 1: Cash economy (Australia)

Risk treatment

Background	A major form of non-compliance is failure to declare cash income. Those involved in cash economy activities: 1) may attempt to remain completely outside the tax system and not register; 2) those who do register, may fail to file or file late, may not accurately report all cash transactions, and may fail to pay or pay on time. Those involved may also be non-compliant with other regulatory obligations such as payment of superannuation guarantee contributions for employees.
Objectives	• A demonstrated ability to detect and respond to the cash economy. • Sustained voluntary compliance by participating taxpayers. • The community is appropriately educated, reducing the tolerance of participation in the cash economy. • Community confidence in the integrity & fairness of the tax system is maintained.
Target group	Historically, the ATO's cash economy risk treatment strategies targeted taxpayers in 'cash economy industries'. These were industries where there was historical evidence of poor tax compliance & a high incidence of cash transactions. Key cash economy industries included: cafes, restaurants and takeaways; cleaning services; hairdressing and beauty; property, building and construction; road freight; smash repairs; taxi services; clothing and textiles; computers; fishing; fruit and vegetables; liquor wholesaling and manufacturing; pubs, clubs and taverns; scrap metal; security services; gold bullion and antique/art dealing; financial services, and tourism and hospitality. The ATO's current cash economy treatment strategies target small business taxpayers exhibiting a range of cash economy behaviours (for example conspicuous consumption and reporting unrealistic income) across industries. Treatment strategies also target small business taxpayers in industries with significant levels of cash transactions, with additional attention given to emerging risks as required.
Risks treated	Principally, failure to accurately report liabilities and failure to file on time
Key elements of risk treatment strategy	The ATO is addressing the risks with a range of integrated treatment strategies. **Communication and assistance:** A variety of information channels such as letters, seminars, internet, brochures and pamphlets, community consultation and business assistance are utilised to inform and educate the wider community about the: • expanding capability to detect cash economy participation and our increased capacity to treat taxpayer non-compliance; • significant costs the cash economy imposes on the community; • practical support we provide to small businesses to enable them to understand and meet their taxation obligations; and • assistance available to businesses experiencing financial difficulties. **Education:** The ATO is encouraging voluntary compliance through the use of a range of educational products designed to support business in meeting their tax obligations. These include tailored letter programs encouraging self-correction, calculators, input and performance benchmarks, seminars and advisory visits. We have the flexibility of utilising these products separately, or as part of our Small Business Assistance Program, to provide targeted information and assistance to cash economy participants. **Engagement:** The ATO works collaboratively with key stakeholders such as tax practitioners and prominent industry/trade associations to improve understanding of the cash economy, and to develop tools to enable them to effectively assess the compliance risk of their member/client base and optimise compliance in areas where the cash economy presents an ongoing risk. **Enforcement:** A suite of differentiated tools and strategies has been developed to treat taxpayers who deliberately or persistently fail to comply with their taxation obligations by participating in the cash economy. These include: • extensive data matching & an automated risk-based case selection model; • a range of targeted audit products (for example risk reviews, correspondence audits, record keeping audits, single risk audits & cash economy audits);

	• a coordinated and comprehensive treatment program to address the cash economy risk in regional communities; • a range of administrative penalties and a number of referrals for prosecution.

Evaluation

Description of evaluation strategy approach	**Indicators being used to evaluate the effectiveness/ impacts of treatments** • Changes to key tax performance indicators at individual taxpayer & industry levels. • Changes in filing performance for activity statements of cash economy participants. • Payment of outstanding liabilities by those in the cash economy. • Trend in the numbers of reports received on the tax evasion hotline. • Trend in the number & nature of visits to web pages relating to the cash economy. • Responses to our surveys, including community, business and tax practitioner perception surveys. • Frequency and tone of media comment relating to the cash economy. • Levels of industry engagement in the management of the cash economy. **What factors may affect the indicators?** • Measuring the effect of compliance strategies & tax performance relies solely on preliminary trend data in the short term. To identify sustained changes in compliance behaviour the ATO will need to observe a number of activity statement periods and income tax annual cycles. • The inability to isolate changes in compliance behaviour that result directly from cash economy strategies. Where possible, external economic influences are considered in the analysis & changes that can be attributed to the surrounding environment. • Media attention may impact compliance behaviour & community perceptions.
Observed impacts of strategy	Effectiveness measurement has been completed on a number of cash economy activities. The following highlights are evident: • Measures of GST/ VAT and income tax performance show that treated (outcome) taxpayers outperform the control population after treatment ends. This improvement is strongly evident for six months after audit and present at a diminishing rate for a further 9-12 months. Treatment does have an impact on Business Activity Statement (BAS) and Income Tax return filing and the improvement in BAS filing is significant. Treatment obviously increases the amount of debt owing, however treated taxpayers reduced their debt after treatment more consistently than comparative groups. • Letters can be effective in generating improved compliance with those having been contacted showing noticeable increases in Income Tax Return and Business Activity Statement self amendments. • Engaging with tax practitioners has proven successful in developing awareness of our overall Cash Economy approach. While measuring the effectiveness of the ATO's response to the cash economy is still at an early stage, initial indications are that the strategies employed are having an impact on the compliance behaviour of those subject to compliance intervention, and community confidence in ATO management of cash economy issues is increasing. There has been a substantial improvement in cash economy reports to the Tax Evasion Hotline and improved community confidence in our management of the cash economy indicated through perception surveys and the level of collaboration between us and industry.

Case study 2: Service entities (Australia)

Risk treatment

Background	The use of service entity arrangements (usually in the form of a trust) lacking an objective commercial basis was widespread, particularly in the legal and accounting professions, but also common among other professions, including architects, engineers and medical practices. Through the payment of excessive or non-commercial fees for the provision of services by the service entity (such as premises and staff), the arrangements allow the

	shifting of income from the professional to the associated service entity, which then streams the income to lower taxed entities (such as family members). The alienation of income from the professional business to associated persons or corporate beneficiaries, results in a lower overall rate of tax payable. A Tax Office review in 2003 revealed that this practice was based upon a broad interpretation of findings in the *Phillips* case, supported by a perception that it had the tacit approval of the Tax Office This perception persisted despite the Commissioner noting the ATO position in the 2001 Annual report and supporting that note in subsequent speeches in 2002. In mid 2005, the ATO published its preliminary views on the validity of these arrangements and after significant consultation with industry the ATO published its final position in April 2006 detailing its views on the information required to support the charges for functions performed by the service entity. The ATO position generated a significant degree of concern for members of the professions and their advisers, and negative feedback for Tax Office.
Objectives	The ATO aims to encourage voluntary compliance with the law by ensuring service entity arrangements are not used incorrectly as a way of minimising the tax payable. By addressing the risk, its desired outcomes were: • Taxpayers understand and comply with the law, and act consistently with Tax Office advice and guidance so that service entity arrangements are commercially defensible and fees are correctly calculated • The professional community has confidence that the Tax Office is appropriately clarifying the law in connection with service entity arrangements and ensuring compliance with the law.
Target group	These were taxpayers with service entity arrangements, professional representative bodies (as intermediaries), and tax practitioners (as intermediaries). The target groups were ranged across the range of professions in the market. Although the range of professions is wide, attention in the first instance was focussed on accountants, lawyers, and general practitioners.
Risks treated	Principally, failure to correctly report tax liabilities and failure to file on time
Key elements of risk treatment strategy	**Marketing and education** • Developed the service entities ruling and explanatory booklet in collaboration with the accounting and legal professions, and professional representative bodies. • Allowed a transitional period for taxpayers (where the case was not high risk) to review their arrangements and align them where necessary to the Tax Office position. This both minimised the risk of them being audited and provided material which assisted them in complying with the law. • Continuing communication and education activities based on feedback and enquiries, including scripts and escalation processes for call centres, maintaining regular contact with legal and accounting firms, providing assistance to tax practitioners with health sector clients, and developing a series of online frequently asked questions and answers. • Managing the media & in particular addressing any misinterpretation immediately. • Using targeted information initiatives including tax practitioner mail-outs, taxpayer and tax practitioner visits, seminars, speeches, updates for National Tax Liaison Group and ATO Tax Practitioner Forum meetings, media releases, surveys and articles. **Tax technical** • Issued a public ruling – Taxation Ruling 2006/2. • Published frequently asked questions and answers (FAQs) online. • Responding to private binding ruling requests & other requests for advice, including meetings with taxpayers & tax practitioners to explain the ATO view. **Active compliance** • Initially auditing high risk cases while providing an opportunity for taxpayers with less aggressive arrangements • Developing criteria for identifying high risk cases & reviewing selected cases. • Monitoring tax performance data to identify cases for review where there is evidence to suggest our guidance material has not been effective in improving voluntary compliance.

Evaluation

Description of evaluation strategy approach	To enable evaluation of the treatments the following indicators were identified: • Trend over time (compared to 2003 baseline) in the number of high risk arrangements identified in the defined population of accountants and lawyers. Success will be demonstrated by a reduction in this number. • Trend over time (compared to 2003 baseline) in the ratio of profit to expenses in service entities in the defined population. Success will be demonstrated by a reduction in this ratio. • Trend over time (compared to 2003 baseline) in the cost base of service entities for the defined population. Success will be demonstrated by only marginal cost base variations over time. • Change in the level of profit disclosed by firms in the defined population, with a corresponding change in the profits of their service entities. Success will be demonstrated by a sustained increase in firm profits coupled with a sustained decrease in service entity profits for the same period. • Trend in the level of queries and requests for advice on service arrangements. Success will be demonstrated by a decline over time. • Trend in the nature of queries on service arrangements. Success will be demonstrated by queries becoming more specific rather than general in nature. • Trend in the level of requests for binding advice. Success will be demonstrated by the number of requests remaining low. • Trend in the nature of feedback into the office from professionals, intermediaries and the Inspector-General of Taxation (Inspector-General). • Trend in the alignment of external communications from professional media with the ATO view. • Trend in the alignment of external communications in popular media with the ATO view. • Trend in the nature of feedback from the professional community. ***Range of factors which may affect the indicators*** Full year tax return data for the 2007–08 income year following the end of the transitional period on 30 April 2007, were not be available until 2009. The ATO used preliminary trend data in the first instance to test the effect of its strategies on compliance and tax performance. In addition, the ATO has difficulty identifying linkages between some professional firms and their service entities. There are also data integrity issues, particularly outside the defined population. Finally, the release of FAQs in May 2008 led to a spike in queries as well as professional and media attention.
Specialised support	The ATO relied upon its normal data matching capability, through systems such as Data Warehouse etc. to link professional firms with service entities.
Observed impacts of strategy	Overall, the ATO has seen a positive trend in the understanding and use of service entity arrangements in line with its guidelines by members of the accountancy and legal professions targeted by these strategies. It also appears there has been a ripple effect across other professions, although the extent of this is yet to be quantified. There has also been positive media comment and support from these professions for the Tax Office position and approach. The ATO has observed some movement from service entity arrangements to alternative structures that continue to alienate business or professional income. It is exploring the means by which this occurs and the legality of the structures and mechanisms now being employed. This trend, if significant, might suggest that while the strategy has been successful in addressing the non-compliant use of service entities, it has not successfully addressed the underlying drivers of the problem and, as a result, the non-compliant behaviours are now manifesting in a different area.
Other comments reflecting on the results achieved	The ATO continues to monitor the profit levels and incomes returned by professional firms and all associated entities including partners in the professional firms. It is taking action to clarify the law regarding potentially aggressive alternative arrangements and structures. The ATO will address the risks currently manifesting with professional firms' compliance, including better understanding the underlying drivers.

Case study 3: Employer obligations (Australia)

Risk treatment

Background	Employers who fail to correctly withhold, report and remit PAYG withholding tax, and comply with their superannuation guarantee obligations, as required by law, have the potential to erode the revenue base and undermine the community's confidence in the tax and retirement income systems. There was evidence of a large pool of employers who had withheld PAYG withholding but had failed to remit the withheld amounts to the ATO. The ATO detected these employers by matching the amounts claimed as credit in their employees' income tax returns. A number of initial assumptions shaped its original compliance strategies in 2004. In particular, it considered that: • A lack of understanding or knowledge of employer obligations was likely to be a key factor underlying much of the non-compliance. • Employers who we assist to accurately calculate and report their obligations and bring their reporting up to date by the close of the audit would improve their compliance from that point on. • A significant field audit presence would generate a compliance 'ripple effect' and leverage improved compliance more broadly than the audited clients. These assumptions have since been evaluated and somewhat reshaped in the light of evidence gathered during treatment For example we found that non-compliance was rarely a result of poor understanding or knowledge of how to comply.
Objectives	• To increase community confidence in the integrity of the tax system by optimising the level of voluntary compliance with employers' obligations to: 1) to report and pay the correct amount of tax withheld from employees' salary and wages, by the due date; and 2) pay when due, the correct amount of employer superannuation contribution to nominated employee superannuation funds. • To understand the drivers and causes of non-compliance so that the ATO can identify opportunities for improving its compliance strategies.
Target group	The target group for this risk treatment strategy is all active employers/withholders. The size of this population in 2008-09 was approximately 785,000 employers.
Risks treated	Principally, failure to pay on time and failure to correctly report liabilities
Key elements of risk treatment strategy	• Direct and indirect messages to employers and intermediaries. • Reviewing cases where employers' appear not to have fully reported pay as you go withholding, including contact by phone to establish whether underreporting has occurred. • Undertaking a program of field audits targeting employers who appear to have failed to correctly report or pay the amounts they withheld. • Ensuring that employers visited have their reporting obligations brought up to date at the close of the audit and are in a position to comply with their obligations into the future. This includes making affordable arrangements for repaying amounts owing as a result of the audit. • Recontacting employers where we detect subsequent non-compliance from longitudinal monitoring.

Evaluation

Description of evaluation strategy approach	Specific measures used to evaluate this strategy include: • Analysis of quantitative and qualitative intelligence on the labour market and on the employers contacted through our field visits collected within the Employer Obligations Intelligence System. • Longitudinal monitoring to measure the ongoing compliance of those employers contacted through our field visits • Analysis of data matching (internal and external sources). • Review of communication and education products available for employers.
Specialised support used	These were: 1) ATO Client accounting systems; 2) Case Management systems; 3) Operational Analytical models & risk engines; 4) audit case selection & compliance activity;

	5) Data warehouse support; 6) Internet & internal internet systems; and 7) Microsoft Access & Outlook.
Observed impacts of strategy	Employer obligations have been undertaking field activities, (auditing between 0.3% and 0.6% of the employer population), since 2004. During this time, the number of employers each year who have paid less than their employees have claimed on their income tax returns remains similar. However, over 90% of employers we audit do not repeat their non-compliance in the following year. Follow-up visits to the 9.5% of the audited population that did fail to comply again in the year following the audit, showed that for some employers a single intervention is not sufficient motivation to change their behaviour. Some will require further monitoring and reminding. In addition to incorrect reporting the study population typically lodges considerably later than the general population. Longitudinal monitoring shows that after an audit, 60% of audited employers improved their filing timeliness. The median lateness for activity statement filing by the audited population (including those that did not improve) reduced from 57 days to 29 days, still well behind the general population whose median filing date is closer to 14 days late.
Other comments reflecting on the results achieved and/or lessons learned	Evaluation of the ATO's strategic approach to the risk has reshaped a number of the original assumptions regarding employer behaviour and what drives it. Observations include: • Approximately 50% of employers who are not complying with PAYG withholding will also fail to meet superannuation obligations in relation to their workers (and sometimes themselves). • Lack of understanding of PAYGW obligations is a very small factor in non-compliance. Only 6% of audited employers had an inadequate understanding of PAYG withholding. Only new employers benefited from education campaigns. • Poor business practices, particularly poor record keeping and inability to use business accounting software, is a major factor in non-compliance. This has wider impact on voluntary compliance than just employer obligations. • Some businesses are only marginally viable, with a significant number technically insolvent. Cash flow difficulties result in amounts withheld for income tax and superannuation contributions being used to maintain business operations. • An active compliance intervention strategy plays a significant part in promoting compliance, but, by itself, is not effective in changing compliance behaviour beyond the audited employers. The ATO has used these insights into the effectiveness of its intervention strategies to reshape its approaches – 'Employer Obligations' is now adopting a more holistic strategy to achieve systemic changes in compliance behaviour. This will encompass directed education, a focus on new business entrants via the Small Business Assistance Program, engagement with advisers and other key leverage points in order to co-design support products and market the approach.

Case study 4: Filing of high tax level (i.e. tax level 6) personal income tax returns (Australia)

Risk treatment

Background	When taxpayers file their returns late it delays revenue collection and puts at risk community confidence in the ATO's administration of the tax system. 'Tax level 6' is the classifier applied to individuals with payable income tax of $20,000 or more in their previous year's assessment. These taxpayers are a key group as their filing compliance is essential to the integrity of the tax system and the community's confidence in the ATO's administration of it. The ATO identified almost 50% of tax level 6 clients as having failed to lodge their 2003 income tax returns on time.
Objectives	To optimise voluntary compliance of these taxpayers who use a tax agent by ensuring they meet their income tax return filing obligations. • Tax agents understand and have the skills to use the Tax Agent Portal to focus on priorities in the return filing program. • Greater certainty among agents that the ATO will follow through on activities it says it will do. • Sustainable improvement in the filing behaviour of tax level 6 taxpayers, evidenced by: 1) improved lodged on-time performance (that is, by 31 March); and 2) an

	increase in the % who file by 30 June of the same filing year for those taxpayers who do not file on time. • A flow-on improvement in the wider community from the increased awareness of the ATO's focus on filing.
Target group	Participants—tax level 6 individual clients of tax agents; tax agents as a group.
Risks treated	Failure to file on time
Key elements of risk treatment strategy	The vast majority of tax level 6 clients engage a tax agent to prepare & file their income tax returns. As a result, the ATO's strategic approach focused on engagement & collaboration with industry representatives to address the risk. The key tactical approaches are described below: • Consulted with the Lodgment Working Group to understand the drivers for tax agents and taxpayers and to co-design and test our approaches. • Helped tax agents to understand the filing program priorities and the system support available. • Advised tax agents of our approach to non-compliance – a preventive approach. • Supported tax agents by communicating the ATO's expectations and planned responses. This would help agents get the information they need in a timely manner. • Ensured ATO's responses to non-compliance were timely and promoted voluntary compliance. • Undertook specific compliance activities: 1) published articles in the TAXAGENT magazine; 2) sent filing reminder letters to tax agents and their tax level 6 clients; 3) initiated tax practitioner broadcasts; 4) issued final notices shortly after the filing due date; and 5) implemented escalated filing compliance activities, incl. prosecutions.

Evaluation

Description of evaluation strategy approach	To enable evaluation of the treatments the following indicators were identified: • Trend in the % of returns filed by the 31 March due date for the tax level 6 population. • Trend in the % of repeat return filings by the due date by the tax level 6 population. • Trend in the % of returns filed by 30 June of the filing year for the tax level 6 population. • Trend in the level of awareness of obligations within the tax agent community. • Trend in the level of awareness of obligations within the targeted tax level 6 community.
Observed impacts of strategy	Tax level 6 returns filed on time for the 2004 year were at 68%. They increased by 17 per cent (to 85 per cent) for the 2005 year, with a further three per cent increase to 88 per cent for the 2006 year and an increase of a further 2 per cent to 90% was achieved for the 2007 year . Of those tax level six taxpayers who filed after the due date, an increased percentage filed within the financial year, rising by four per cent to 95 per cent for the 2005 year and a further rise of one per cent to 96 per cent for the 2006 year, which was maintained for the 2007 year. Tax agents with tax level six clients generally have improved on-time return filing performance across their whole client base. Feedback from the Lodgment Working Group indicates increased awareness among practitioners and their clients. "Tax level 6 is viewed as a brand" according to one tax agent.
Other comments reflecting on the results achieved and lessons learned	The key learning for the ATO in its approach to strategy development was that engagement with tax agents, as the key leverage point for tax level 6 taxpayers, allowed it to identify underlying causes of non-compliant behaviour and work with agents to co-design the strategy to effect the desired change. The ATO learned that the tax level 6 taxpayer base is a highly variable population, and this would appear to confirm that its strategies have primarily affected tax agents rather than the taxpayers themselves. It has also found improved filing performance across the entire clientele of those tax agents dealt with regarding their tax level 6 clients.

Case study 5: Atlantic region**-** Underground Economy Compliance Measurement Initiative (UECMI) covering the construction industry (*pilot project*) (Canada)

Risk treatment

Background	Historically, program funding for the National Underground Economy Initiative is based on CRA's ability to deliver mandated outputs. These mandated outputs translate into budgeted outputs across CRA, such as number of audits, revenue generated from audits, number of community visits, etc. CRA develops strategies to meet those budgeted outputs, and captures data to report on the production of those outputs. This production data speaks to ongoing annual results, but does not allow CRA to comment on the impact it is having on future compliance. Under this initiative, the Atlantic Region is not required to meet traditional output targets. Rather, the Region is required to pilot alternative approaches to addressing the underground economy, and measure the impact that these approaches have on future taxpayer compliance. The project will use a basket of strategies aimed at both the supply and demand sides of the ***construction industry*** over the three phases of the project. Non-compliance in this industry has been well documented over the past few decades. Each phase will target a specific aspect of the UE in the construction industry and will have appropriate evaluation measures tailored around each strategy.
Objectives	The primary objectives of the Atlantic Region UECMI are to: 1) pilot alternative strategies aimed at having an impact on improving future compliance in the construction industry; and 2) appropriately measure the impact of these various strategies.
Target group	The target group for the supply-side of the industry are registered operators within seven sub-sectors of the construction industry in Atlantic Canada (approximately 4,700). The target group for the demand-side of the industry are Atlantic Canadian homeowners (approximately 900,000).
Risks treated	In priority order—failure to correctly report liabilities, pay on time and to file on time.
Key elements of risk treatment strategy	*Phase 1 (fiscal 2008/09)* ***Books and Records Reviews:*** Letters were mailed to 3,700 industry participants, chosen randomly from a population of 4,700. Attempts were made to conduct Books and Records reviews on 2,000 accounts out of this population. These reviews ensure entities operating in the construction industry maintain adequate records for CRA purposes. Underground economy issues were discussed with the taxpayers. The objective of these discussions was to gather information that would aid in developing strategies to address the underground economy in the future. ***Industry Consultations:*** Letters were mailed to selected trade associations to introduce the compliance strategy and to provide industries the opportunity to meet with CRA. CRA drew on the industry knowledge to help identify factors that drive the underground economy in the construction industry, and will work with these associations to develop partnerships and discuss strategies to further combat underground activity in the region. *Phase 2 (fiscal 2009/10)* ***Address Enforcement of the Construction Payment Reporting System (CPRS):*** Underground economy resources in Audit will increase the coverage of monitoring the CPRS, on both the 'payer' and 'recipient' sides. The goal of this strategy is to measure the impact on future CPRS compliance behaviour through a concentrated education and enforcement campaign across Atlantic businesses in the construction industry: • Audits restricted to labour expenses are conducted on a population of corporate taxpayers (payer-side) where there are indications of potential non-compliance within CPRS; • Audits restricted to revenue from third party information (the Construction Summary Worksheet Program – GST/HST Housing Rebate) are conducted on a population of individual taxpayers; • Taxpayers will be educated on the importance of CPRS compliance. *Phase 3 (fiscal 2010/11)* ***Communications Strategy:*** Proactive and innovative strategies will be implemented to impact the attitudes and behaviour of the homeowner to prevent or decrease the likelihood of consumer participation in the underground economy: • Just-in-Time communication is delivered to region homeowners; • Stakeholder partnerships for the homeowners' strategy are pursued.

Evaluation

Description of evaluation strategy approach	Each phase described above is utilizing an experimental or quasi-experimental program design to assist in determining if there is a causal link between the activity and resulting changes in behaviour/attitude. The key behaviour indicators are: returns filed; number of days returns filed late; amount of taxes reported; amount taxes due after assessment. Other potential causes to behavioural change, such as economic and demographic factors are taken into account through the experimental methodology. Randomization balances out these other factors across the groups, leaving only one difference between the two groups; the group it was randomly assigned to (e.g. Treatment vs Control). The control group acts as "what would have happened if we had done nothing". The indicators were developed to determine the impact that a program had on filing, reporting and paying behaviour, not compliance specifically. Using amounts reported as an example; if the median amounts (or transformed means) reported in the treated group are larger than in the control group, and the difference is deemed to be statistically significant, the experimental design allows this difference to attributed back to the treatment. The CRA will be in a position to infer that treatment A increased/decreased taxes reported by $X (within a certain degree of confidence), but not how much is unreported.
Specialised support	This project has required an understanding of social science methodologies, including experimental and quasi-experimental design, as well as an understanding of various categorical and continuous approaches to data analysis (ANOVA, logistic regression). Specialized statistical software (eg. SAS, SPSS, etc) is being used.
Observed impacts of strategy	Phase 1 field work was completed in March of 2009 and the first behavioural indicators (filing, reporting, and paying) were reported in the summer of 2009. Data collection is ongoing. Phase 2 of field work will be completed in March of 2010. Request for data will be submitted in the summer of 2010. Phase 3 field work will be completed in the fall of 2010. Pre-program surveys will be delivered in February/March 2010. Post-program surveys will be delivered immediately after field work.
Other comments reflecting on the results achieved and lessons learned	• Approaches to measuring impacts require a strong understanding of social science methodologies and accompanying resources which are scarce in most tax administrations. • Experimental approaches are strongest at being able to discount other possible reasons for the observed behavioural/attitudinal changes and should be used wherever possible. • Communication with the field auditors and managers is necessary if "measuring outcomes" is not business-as-usual. • Senior management at all levels should be engaged from the outset. • Approach to randomization can be complex depending on population (e.g. one individual with multiple businesses in target sector, multiple individuals attached to one business in sector). • Define the target population based on how field auditors'/managers' view of the population; this will assist with ease of information dissemination. • Use simple, straightforward experimental approaches at first, ideally ones that will allow comparisons of current treatments and new treatments.

Case study 6: Employees tips and gratuities incomes (*pilot project*) (Canada)

Risk treatment

Background	The risk treatment was developed to measure the success of the Tips and Gratuities Pilot Project. The purpose of the pilot is to promote compliance in the reporting of tips and gratuities in the hospitality industry, through education and outreach, followed up by verification activities.
Objectives	The objective of the risk treatment is to measure the level of compliance prior to the project, during the project period and after the project to determine the effects of our efforts.
Target group	The target group for this risk treatment strategy is all employees of restaurants, and wineries, golf clubs and hotels with restaurants in the Niagara Region with sales in excess of a certain threshold (sales in excess of $250,000). In terms of size, letters were sent to 450 establishments in the Niagara Region, which is a very concentrated area in Ontario.
Risks treated	Failure to correctly report liabilities in returns filed and failure to file returns on time.

Key elements of risk treatment strategy	The key elements of the risk strategy are: • CRA news release and letters to the establishments. • Presentations to owners and/or servers at establishments with the goal of 100% coverage. This includes the distribution of pamphlet "Talking About Tips and Gratuities", which was developed for the purposes of this project. • Collaboration with external partners such as: 1) Niagara College; 2) Restaurant Association and Chamber of Commerce for local municipalities; & 3) Smart Serve providers, Service Canada, & other employment centres. • Selection of establishments for compliance action (6 months after presentation) • Reassessment of employee servers (and establishments where necessary) • Forwarding T1 Adjustment Forms (to request changes to years outside of audit period) to employee servers with Final Letter.

Evaluation

Description of evaluation strategy approach	The CRA will measure the amount of education and outreach performed in the community by comparing the number of outreach presentations against known establishments. The CRA will measure the level of compliance by: • Comparing reported tips for years 2005 to 2009 to enable it to compare the level of compliance before and after the project—by obtaining a sample of establishments where it made an Outreach presentation, it will measure the reported Tips as a percentage of T4 Income (wages). It will compare that percentage before and after the Outreach presentation date. • Measuring reassessments from audit action. • Measuring the number of T1 adjustments filed whereby taxpayers report tip income for years outside of the audit period—by measuring the number of T1 Adjustment Requests (specifically requests to add Other Income/Tip Income) made by employees that have been subject to audit action (reassessed), received subsequent to our audit. • By measuring the reported Tips as a percentage of T4 Income (wages) for establishments under Audit action, before and after the audit (this information may not be available until the fiscal after the audit concludes in some cases). • Measuring the number of Voluntary Disclosures relating to unreported tip income.
Specialised support	• Identification of establishments through NAICS (industry) codes and local knowledge. • Assistance from Electronic Commerce Audit Specialists to obtain electronic books and records as well as to obtain and verify electronic point of sale information from establishments under audit. By using electronic data, actual tip information with respect to debit cards and credit cards is obtained from the establishment's point of sale (POS) system specific to each server. Additionally, cash sale information is obtained from the POS system, to which a conservative percentage is applied to estimate cash tips. • Used standardized CRA tools to measure the tax earned by compliance actions & to monitor where taxpayers have requested to report additional tip income outside of the audit period. • Contact with other areas (Voluntary Disclosures Program).
Observed impacts of strategy	CRA have discovered that: • There is significant risk with respect to tips and gratuities as audit activities have resulted in the discovery of $830K in unreported TIPS over 2 audit years (2 establishments & 59 employees to date). • An additional 9 establishments are in various stages of compliance action. • Nine servers (15% of those reassessed) have filed adjustments to their taxable income for years outside of the audit period, representing an additional $59K in unreported tips. This is directly tied to the project as T1 Adjustment forms were mailed at the conclusion of the audit to each server. • Most individuals have agreed with changes to their taxable income. • In many cases, the establishments deduct money from servers' tips. This has not always been reported, resulting in reassessment of the restaurant. • One referral was made to the Non-Filer area for a server who has not filed tax returns for the years under audit. • Three leads have been made on Sub-Contractors who performed work for the

	establishments. It does not appear that this income was reported.
Other comments reflecting on the results achieved	• Outreach presentations have been made to 278 establishments (87%) to date with the objective of obtaining 100% coverage by March 2010. • There has been significant amount of unreported tips in the audits of establishments completed to date: $830K in unreported tips from the 59 servers and $197K in unreported income from the 2 establishments (money deducted from the servers' tips). • Additionally, 9 servers have requested to add tip income to years outside of the audit period totalling $59K. • The use of electronic data has enabled the reassessments to be prepared efficiently. The servers are frequently concurring with the results as it is difficult to dispute the electronic data obtained from the establishment. • CRA considers that its efforts are having an educational impact on the industry's culture. Industry insiders often tell servers that they only need to report 10% of their T4 wages as tip income. Results are displaying to the servers that their tips are more likely 100% to 200% of their T4 wages. • Some of the establishments under audit action are higher-end restaurants with large profiles. CRA believes that this will have a significant "word of mouth" effect on the local industry (difficult to measure).

Case study 7: Compliance Indicators for the Trucking Industry (*pilot project*) (Canada)

Risk treatment

Background	The purpose of the pilot is to address suspected Underground Economic activities within the Trucking Industry via typical risk treatments, such as income tax audits, federal and provincial sales tax audits, IFTA (International Fuel Tax Agreement) audits and payroll examinations. The expectation is that these compliance activities will improve short to mid term compliance rates for those individuals subjected to our compliance activities. An analysis using the size of the industry and the known rates of non-compliance as key measures suggested that this industry has a higher absolute volume ($) of non-compliance than many other high-risk industries in Canada.
Objectives	The objective of the risk treatment is to improve filing, reporting and paying compliance in the target groups. This will be done by calculating the difference in the compliance rates between the study and control groups before and after the program period to determine whether or not we have had any success in changing the compliance behaviours of this group.
Target group	The target group for this risk treatment strategy is self-employed, independent and small operators in the Truck Transportation industry in the Ontario and Pacific Regions. In terms of size, the truck transportation industry is quite large and contributes $15 billion per year to national Gross Domestic Product, ranked 25th out of 72 industry sub-sectors.
Risks treated	Failure to file, failure to correctly report liabilities, & failure to pay on time
Key elements of risk treatment strategy	The key elements of the risk strategy are: 1) Audit activities; 2) Trust examinations; and 3) Collaboration with the Province of Ontario's Employer Health Tax (EHT) Audit Section.

Evaluation

Description of evaluation strategy approach	• A sample of operations was randomly selected from a target population. From this sample, CRA then assigned audits/trust examinations to the study group on a stratified random basis. • To be able to measure the impact of its treatments, CRA also identified a comparison group who did not receive the compliance actions. This comparison group was also randomly selected from the sample. • By randomly assigning businesses from the sample to a treatment and control group and making comparisons between the two, the CRA has a program design that gives it an opportunity to determine whether or not changes in measurable behaviours between the two groups were caused by its compliance activity.

	The three indicators being used are: 1) Filing Compliance = $$\frac{\text{Change in \% of Study Group participants filing on time}}{\text{Change in \% of Control Group participants filing on time}}$$ 2) Reporting Compliance = $$\frac{\text{Change in the Average Net Income of Study Group participants}}{\text{Change in the Average Net Income of Control Group participants}}$$ 3) Payment Compliance = $$\frac{\text{Change in difference between Net \& Gross Tax at risk* in Study Group}}{\text{Change in difference between Net \& Gross Tax at risk* in Control Group}}$$ *(* At risk refers to using only those taxes that were not deducted at source as the denominator of the ratio of taxes unpaid. For example, total taxes due $10,000, deducted at source $8,000, paid upon filing $1,500. Although 95% of the total tax bill was paid, only 75% of the tax bill that was in the control of the taxpayer was actually paid. In terms of timing, this refers to "upon assessment").* The filing and payment compliance indicators are direct and complete measures of compliance: does our treatment group now file on a more timely basis and pay a higher % of their liabilities on a timely basis than our control group? The reporting compliance indicator, however, is only an indicator of how much was reported, not how much should have been reported. This speaks more to reporting behaviour than reporting compliance.
Specialised support	This has included an understanding of social science methodologies, including experimental and quasi-experimental design, as well as an understanding of various categorical and continuous approaches to data analysis (ANOVA, logistic regression) are required. Specialized statistical software (e.g. SAS, SPSS, etc) is highly recommended
Observed impacts of strategy	• Phase 1 (CRA audits/Trust Examinations) were completed in the spring of 2009. The analysis of the impact that the trust examinations had on compliance in currently underway. Data collection is ongoing for the remainder of the revenue streams (T1, T2, GST/HST). • Phase 2 (Ontario EHT audits) will be completed by February 2010. Data collection will begin at that time.
Other comments reflecting on the results achieved	• Approaches to measuring impacts require a strong understanding of social science methodologies and accompanying resources which are scarce in most tax administrations. • Gaining a better understanding of compliance behaviour is a complex and long-term undertaking; patience is required. • Experimental approaches are strongest at being able to discount other possible reasons for the observed behavioural/attitudinal changes and should be used wherever possible. • Communication with the field auditors and managers is necessary if "measuring outcomes" is not business-as-usual. • Senior management should be engaged at the earliest stages of project. • Approach to randomization can be complex depending on population (e.g. One individual with multiple businesses in target sector, multiple individuals attached to one business in sector). • Defining the target population on how field auditors/managers view the population will assist with ease of information dissemination. • Use simple, straightforward experimental approaches at first, ideally ones that will allow comparisons of current treatments and new treatments.

Case study 8: Information and guidance in a shareholder taxation project (Denmark)

Risk treatment

Background	New tax legislation came into force in January 2006 involving new regulations for the taxation of profits on shares and other securities. As many people find the rules for taxation very complex, it was assumed that the level of compliance was low in this area. The risk

	treatment was a nation-wide initiative in the years 2007 and 2008.
Objectives	The objective was to increase the overall compliance in relation to ownership of shares by increasing: • Awareness and ability – by providing information and guidance while at the same time lowering barriers to compliance by offering a new online computer programme • Willingness – by pointing out the fact that SKAT was aware that the recipient of the guidance material had sold shares or securities worth a substantial amount.
Target group	A target group was selected from among wage-earners and other similar tax-payers who owned shares or investment securities. There were a number of criteria for inclusion in the target group, including holdings in a particular type of investment company and sale of shares or securities for at least DKK 100,000 (roughly equivalent to €15,000). This filtering procedure was applied to all the taxpayers in Denmark, and resulted in the identification of a target group of 79,107 people. Though the initiative was targeted at this specific group, the target group was potentially much larger and contains all taxpayers who own shares or securities.
Risks treated	Failure to correctly report liabilities
Key elements of risk treatment strategy	An information leaflet targeted at ordinary taxpayers was produced and sent out to around 5,000 taxpayers selected at random from the target group. A guidance letter was accompanying the leaflet providing guidance on what to do and where to go for help and at the same time drawing attention to the fact that SKAT was aware that the recipient had sold shares or securities worth a substantial amount. In addition, an effort has been made to counter the problems many taxpayers experience in following the complex regulations through the use of a new online computer programme allowing taxpayers to access electronic assistance with the complex calculations. Finally, in the second half of 2008 a large number of checks were carried out in order to make it evident that attempts at cheating in this area did not pay.

Evaluation

Description of evaluation strategy approach	Audits were carried out for the recipients of the guidance material and for an equivalent group who had not received the material, again selected at random from the target group. Since both groups were large in size and were randomly selected from the same population, the effect of the guidance material was calculated quite simply as the difference between the percentages of cases in which adjustments were made to taxable income and between the amounts involved.
Observed impacts of strategy	In general the project revealed that the extent of errors in declaration of income from shares and securities was quite considerable and that the initiative was especially effective at the more straightforward end of the spectrum of the selected group of taxpayers who had share income. The guidance material significantly reduced the percentage of cases where errors had been made by over six percentages points, from approximately 58 to 52 percent. Considering the overall average amount of adjustment to taxable income, the amount was larger for those who had not received guidance than for those who had. However, the difference was only significant for those cases where assessed income from shares was reduced. The average net adjustment per case was DKK 14,000 without guidance and DKK 12,000 with guidance. The results are based on checks made on the first one-third of the 10,000 taxpayers in the random sample. This one-third was again selected at random from the sample to ensure that the results are representative of the target group. The checks were carried out in the second half of 2008.
Other comments reflecting on the results achieved	***Though the guidance initiative had a significant effect, the analysis also showed that it was not possible to effectively remove the problem by using known treatments. Third party reporting facilitating the taxation of shareholdings was, therefore, introduced as part of a major tax reform in January 2010.*** These changes in the requirements of third party information from financial institutions are designed to address the low compliance levels documented by this project and by a compliance survey based on random audits. The risk has thus now been addressed in a cost effectively manner as a consequence of the two analysis. The use of drawing random samples from a random sample when planning which audits to do first has been very useful in order to draw early conclusions and in order to adjust a given treatment. This way of planning enables us to avoid the bias in early analysis caused by the fact that the easier cases with low proportions of adjustments generally were finished more quickly.

(For more details, see http://www.itdweb.org/Pages/Search.aspx?lang=3&st=6&sort=2&c=55

Case study 9: Campaign directed at Danish residents who own property overseas (Denmark)

Risk treatment

Background	Danish taxpayers who own property overseas are taxed by approximately 1 percent of the market value of the property. Various regulations and exceptions make the rules complicated. At the same time it could be observed that the number of declarations of property overseas was low relative to what could reasonably be expected. The risk treatment was a nationwide initiative launched in 2007. There is still focus on the area although the risk treatment has been modified slightly in subsequent years
Objectives	The objective was to increase the level of compliance relating to the declaration of property overseas by increasing: • Awareness of the rules in the area – through a nationwide information campaign. • Willingness to comply with the rules – by sending out letters explaining the rules and implicitly making it clear that SKAT was aware of the property in question.
Target group	The primary target group was Danish residents who owned, or were about to purchase, property overseas. A secondary target group consisted of real estate agents, mortgage providers and financial institutions involved in the purchase of properties abroad.
Risks treated	Failure to correctly report liabilities
Key elements of risk treatment strategy	The risk treatment involved a variety of initiatives aimed at increasing awareness and knowledge of the rules, including advertising in relevant media, production of an information leaflet accessible from the tax authority website, visiting estate agents and participating in trade fairs. In addition, letters were sent out to Danish residents owning properties in Sweden. This part of the treatment was conducted on the basis of information received from the Swedish tax authority in connection with a Nordic agreement on exchange of information. The letters explained the rules and implicitly made it clear that the property was not invisible to SKAT.

Evaluation

Description of evaluation strategy approach	The effect of the campaign in general was assessed as the difference between the number of declaration of properties overseas prior to and after the campaign was conducted. This was not an exact measurement of the effect but gave an indication of the effect of the overall campaign. It was, however, possible to measure the effect of the campaign directed specifically at the owners of property in Sweden. A treatment group received the letter requesting them to take note of the information supplied by the Swedish authorities prior to the deadline for declarations for the tax year 2006 whereas the control group did not receive that letter until the deadline had passed. Since the two groups selected from the population of property-owners provided by Sweden can be assumed to represent random samples, the effect of the information letters can be calculated quite simply as the difference between the percentages who declared property overseas in the two groups.
Observed impacts of strategy	***The number of people who declared overseas property doubled for the tax year 2006 compared to 2005***, clearly indicating that the overall campaign successfully has affected behaviour in this area. Although the analysis with great statistical certainty demonstrated that there was a link between the declaration of property overseas and the fact that owners were personally notified that SKAT was in possession of information about their ownership of the property, no clear conclusion can be drawn on how much each of the various elements of the initiative contributed to the overall increase.
Other comments reflecting on the results achieved	The increase in tax revenue that will arise as a result of the campaign is estimated to be approximately DKK 131 million (roughly equivalent to EUR 17.5 million) taking into account only the future revenue from taxes on the value of the property. If secondary effects such as tax on rental income or tax on possible profits from sales of declared properties are also included, the increase in revenue is likely to be significantly greater. As 20-30 FTE were used in the project, there is a clear indication that the campaign has been particularly beneficial in revenue terms. It is also a good example of an area where traditional audits would not be cost-effective, but where sending out information does bear fruit. SKAT is now receiving information from both Swedish & Norwegian revenue bodies according to the Nordic agreement and is currently trying to establish a similar agreement on exchange of information in this area with other European countries.

(For more details, see http://www.itdweb.org/Pages/Search.aspx?lang=3&st=6&sort=2&c=55

Case study 10: A local driving instructor project (Denmark)

Risk treatment

Background	The project was initiated as intelligence indicated low levels of compliance with respect to income tax and VAT among driving instructors, who are normally self-employed and run their own "driving schools". The risk treatment was a small, local initiative in the Danish town of Horsens from March 2007 to February 2008.
Objectives	The objective was to increase compliance with respect to income tax & VAT expressed as declared turnover among the target group. A secondary objective was to gain an overview of compliance levels of the sector as a whole.
Target group	The target group included all 77 driving schools in Horsens. 17 of these were audited.
Risks treated	Failure to correctly report liabilities
Key elements of risk treatment strategy	The project was based on a traditional audit approach, but evolved to include visibility and cooperation components, as an initial press release and subsequent press interest resulted in substantial visibility while contacts with the local driving instructors' association resulted in the participation of tax representatives in a "theme day" for members.

Evaluation

Description of evaluation strategy approach	The evaluation assesses the effects of the project on the driving instructors who were audited and on those who were not. The effects were expected to show as an increase in declared turnover relative to the national average. But because factors *other than* compliance with tax legislation contribute to declared turnover, it is difficult to isolate the effect of the project. Therefore the evaluation use average turnover *per driving test* in order to rule out variations in the level of activity in the business. The resulting measure can be considered *an indicator of effect*.
Observed impacts of strategy	The audits related to the tax year 2005 and led to adjustments to income tax and VAT in 16 out of the 17 cases. The audits generated a net increase of nearly DKK 1.6 million. Even though the high percentage of adjustments was primarily due to the fact that the selection of cases for investigation was not random but based on a risk assessment the result revealed low levels of compliance in the sector. In the case of the instructors who were audited, the effect of the initiative was apparent in a very noticeable increase in the declared amount of turnover from the tax year 2005 to the tax year 2006. There were also signs that the initiative had a longer-term effect, since the level of declared turnovers appeared unchanged between 2006 and 2007. In the case of the instructors who were not audited, an effect was manifested as an increase in the average declared level of turnover per driving test from the tax year 2006 to the tax year 2007 significantly greater than the national average. This indicates that the group in general has been influenced by the "theme day" and the press coverage of the results of the project.
Other comments reflecting on the results achieved	The non-random nature of the audits made it necessary to base the analysis solely on indicators. As such the project provide an example of how less ambitious impact evaluations can be carried out for smaller projects where less resources are available and a lower level of evidence is acceptable. Visibility and cooperation appear to be cost-effective means of affecting compliance that can complement a more traditional enforcement approach based on risk-based audits in order to reach a larger target group. Therefore these components should ideally be considered from the outset of a project.

(For more details, see http://www.itdweb.org/Pages/Search.aspx?lang=3&st=6&sort=2&c=55

Case study 11: Compliance with requirements to maintain a staff ledger in the hairdressing and restaurant industries, in order to reduce undeclared work (Sweden)

Risk treatment

Background	In 2005, the Swedish government proposed new legislation stipulating that all businesses in the ***restaurant and hairdressing*** industries are obliged to keep a daily staff register/ledger. The ledger should, on a daily basis, include an entry for each employee working on that particular day. The National Tax Agency would be able to conduct unannounced inspections of the ledger as well as limited audits in connection with the visits. If employers fail to put in place a staff register they are charged a fine. The Parliament approved the proposal in June 2006, and the legislation came into effect in January 2007. The background of the legislation was the significant tax evasion in cash industries, in

	restaurant and hairdressing industries. These industries were particularly exposed in terms of undeclared revenue and undeclared work. This was acknowledged both by trade federations and the Tax Agency. The law allows the Tax Agency to conduct unannounced inspections at workplaces to determine if employers are registering their staff in the correct manner. If not, the employer may be charged a fine. The agency could also choose to go ahead with a full in-depth audit if needed. If the workplace does not have a staff register in order, a fine of SEK 10,000 (about €900) can be imposed, along with an additional fine of SEK 2,000 (€180) for each employee who is not registered. The staff register/ledger has to include the name of the business, the employees' names and birthdates and their working hours. Registration in the ledger, takes place when employees begin their shift and ends the shift. The Tax Agency performed 31,000 inspections at workplaces in 2007, 18 000 in 2008 and 17, 000 in 2009, covering around 80% of the total number of businesses.
Objectives	Seeks to reduce and prevent undeclared work
Target group	All businesses in the restaurant and hairdressing industries. In 2007, when the legislation came into effect, there was about 29000 restaurants and 2400 hairdressers with employees.
Risks treated	Failure to correctly report liabilities and failure to register
Key elements of risk treatment strategy	The overall strategy of the Swedish Tax Agency is "Right from the Start", i.e. errors should as much as possible be prevented and treated as early as possible in the process. The entry into the staff ledger takes place <u>before</u> the book-keeping and submission of the tax return. Therefore, the staff ledger could be seen as a preventive measure regarding payroll tax, and indirectly regarding income tax and VAT ('Failure to correctly report liabilities'). Furthermore, the obligation to keep a ledger also give the Tax Agency the ability to more easily detect non-filers while doing unannounced visits ('Failure to register').

Evaluation

Description of evaluation strategy approach	The evaluation had two parts, one where the aim is to measure the impact of the legislation, the second part deals with the effects on employment and the entrepreneur's perception of staff ledgers and the tax control. **The first part (impact):** To measure if staff ledgers have an effect on tax evasion it is not enough that a staff member is recorded in the staff ledger. Information must also be included in the companies' accounts and thereby influence the payroll. If payroll accounts increase due to staff ledgers, earlier undeclared ('black') wages is assumed to be declared ('white'). The evaluation has focused on measuring changes in payroll accounts, which are likely to be caused by the legislation of the staff ledgers. The initial effect is calculated by payroll accounts before the rule change compared to the payroll accounts after the rule change. Since it is a short interval of time would influence from external factors to be almost negligible. In order to calculate the long-term effects it is not sufficient to measure only the change of salary accounts. To get an idea of the impact of cyclical changes had on the payroll accounts, comparisons were made with several different groups of companies and development dimensions. The results are then compared with actual payroll accounts. The evaluation has used seven different calculations for the restaurant and five for the hairdressing industry to estimate the long term impact of staff ledgers. There is uncertainty in each of the estimates made of the effects on tax evasion. The calculations show, however, the same direction - staff ledgers have been effective. **The second part (perception):** The evaluation was carried out by a study of in-depth interviews and focus groups and a survey with telephone interviews (500 restaurants and 500 hairdressers). Some of the interviews were also filmed. In addition, the study addressed a questionnaire to tax officials who worked with inspections at workplace. The purpose of the survey was to reflect the entrepreneurs' perception through the officials companies. Finally, there was a comparison with the results of a survey directed to the restaurants made by a trade publication.
Observed impacts of strategy	**The first part (impact)** ***Staff ledgers have had a significant impact on the reported salary in the accounts in the restaurant industry*** and the effect has increased gradually during the first two years after the legislation came into effect. The uncertainty of the estimates is large. The overall conclusion is that a large amount of undeclared work has become "white" as a result of the staff ledgers. The effect of staff ledgers the first month after the introduction calculated with increase reported earnings SEK 725 million per year (around €80 million) or 3, 100 AWU (annual

	work units, i.e. employees). This represents 6.4 percent of the wage bill for January 2007. The long term impact has resulted in an increase in the reported wages in 2007, located in or near the range of SEK 0.7-1.2 billion (around 70-120 million, or about 3,000-5,000 AWU or an increase of about 5 -8 percent of the reported wage bill in the restaurant industry. The effect has increased further in 2008. Together with effect from 2007 the increase is about the range of SEK 1.0-1.5 billion (around €100- 150 million), or about 4,000-6,000 AWU or 7-11 percent of the reported wage bill in the industry. The effects of staff ledgers of businesses of all sizes are difficult to measure with a higher degree of reliability. The analysis shows that staff ledgers are more effective for private firms and partnerships than for limited companies. Since limited companies have more extensive operations the economic impact of personnel ledgers are still higher for limited companies. ***Staff ledgers have had some effect on the reported salary in the accounts the hairdressing industry*** directly after the legislation came into effect, but the effect has worn off over time. The effect of staff ledgers in the first month after the introduction in 2007 calculated increase reported earnings with SEK 44 million per year or 190 AWU. This represents SEK 4.4 percent of salary bill for January 2007. The calculations show the long term impact of the staff ledgers in the hairdressing industry the reported wages in 2007 increased by approximately SEK 20-60 million, or about 70-250 AWU or an increase of 2-7 percent of the reported salary bill in the industry . The impact has slowed sharply in 2008 and only one third of the first year effect persists. For the year 2008 is the uncertainty so great that it is impossible to determine any effect of staff ledgers. We can not draw any conclusions about the impact of staff ledgers differ between large and small companies within the hairdressing industry. On the other hand, the analysis shows that staff ledgers, both in terms of amount and the relative salary bill is more effective for private firms and partnerships than for limited companies. **The second part (perception)** A majority have a positive perception of staff ledgers. A smaller proportion (i.e. 20 percent) has a negative perception. The most frequently mentioned benefits from the entrepreneurs' responses is that staff ledgers help to "keep tabs" on the staff and their working time as well as to reduce undeclared work in general. There are relatively few, only ten percent, which is opposed to the tax officials can do a inspection without announcement. Most entrepreneurs perceive that the visit by the Tax Agency has been implemented smoothly, respectfully and with a good response. However, there is a minority that is negative to how the tax official responded to them during the inspection. The majority believe that it is easy and it is quick to fill in the staff ledgers. The telephone interviews revealed that the system is not perceived as time-or cost-consuming. However, it is psychologically burdening as it creates anxiety and stress. The entrepreneurs feel that there are many inspections which make entrepreneurship more difficult and less attractive to hold. Entrepreneurs in the focus groups were highly critical of the inspection fine and felt that it was not related to the level of misconduct. The result from the focus groups also showed that entrepreneurs feel that it is intrusive to focus on only two sectors, which could lead to a general public perception that it is these industries that have the most undeclared work.
Other comments reflecting on the results achieved	After the evaluation a public commission has in late 2009 submitted proposals for extending the legislation of staff ledgers to the construction and laundry industries. The experiences of the working methods were used in the introduction of a similar but larger preventive measure initiative, certified cash registers (introduced 1 January 2010). The evaluation was extensive, time consuming and demanded a high level of expertise, e.g. highly qualified statisticians and economists. This type of evaluation could therefore only be performed in a few highly prioritised cases. For the staff ledgers, there was an extensive political demand as the results was to be used in the decision making process of a possible extension. A challenge in the evaluation was to isolate external factors such as the general economic development. It was also difficult to differentiate the impact resulting from the visits and the introduction of the legislation itself. The results of the evaluation, i.e. number of 'black' jobs turned 'white' and increased tax collection, were easily communicated and understandable for the public and received fairly extensive media coverage. The perception of the traders and the filmed interviews are being used in the training programme of the Tax Agency regarding staff' attitudes towards taxpayers.

Case study 12: Post- compliance monitoring/measurement of real estate agent industry (Singapore)

Risk treatment

Background	IRAS shares with the public its compliance focus and plans on a yearly basis as it believes that this will contribute to its overall strategic outcome to enhance voluntary compliance. The real estate agent industry was identified as one for compliance focus in the year 2007 due to the active property market in that year which resulted in high volumes of property transactions. As the risk of under-reporting and overstatement of expenses heightened during this property sector boom, this industry was highlighted as a compliance focus for that year. During the audit case selection process, IRAS approach was made more accurate with the collation of third party information provided by the real estate commission paying companies in the form of commission listings. With this reliable third party information on hand, IRAS was able to verify the agent's reported commission income and observe if the reported adjusted profit ratio fell within the industry average. IRAS also took conscientious efforts to publicise any post compliance monitoring results in the following year's compliance publicity plans in order to close the loop by sharing with the public the results of our compliance efforts.
Objectives	The intended outcome sought is to prevent, detect and deter non-compliance in this industry. Through the selection of specific high risk cases for audit review, we hope to detect cases of non-compliance and thereafter to reform their behaviour.
Target group	The target group of taxpayers for this risk treatment strategy is the real estate agents and there are approximately 20,000 real estate agents in Singapore.
Risks treated	Failure to report, failure to file on time
Key elements of risk treatment strategy	High risk cases were identified for audit review. Some indicators of high risk were a discrepancy between taxpayers' declarations and external information from commission paying agencies as well as financial ratios that are outside of the industry norm. In order to ensure that IRAS's compliance program reaches a larger group of taxpayers besides those selected for audit, educational seminars were carried out on this industry to further educate agents on proper record keeping and the correct claiming of expenses.

Evaluation

Description of evaluation strategy approach	A 3 year monitoring of the compliance level of the general population as well as the pool of audited cases amongst the real estate agent industry was to be carried out to measure the effectiveness of IRAS's compliance program on the industry as a whole. Some indicators of compliance level that are being monitored include filing rate and financial indicators such as adjusted profit ratios. For selected cases that meet IRAS's criteria, it will also carry out post compliance monitoring on them over the following 3 years.
Specialised support used	In order to take into account any external factors that might affect the financial indicators being monitored, economic factors such as the overall change in the prices of properties were considered. These external factors were obtained from external sources such as the statistical series published by URA (Urban Redevelopment Authority of Singapore).
Observed impacts of strategy	There was an overall improvement in the financial indicators of both the audited pool and general population over a time frame of 2 years after the end of our audit program. This is despite taking into account the spike in the property prices over the last 2 years which may have contributed to the better financial indicators. Specific cases that have been selected for post compliance monitoring has also shown satisfactory performance in their onward reporting.
Other comments reflecting on the results achieved	IRAS will consider selecting a control group (whereby no audit actions are taken) to closely monitor their compliance level to further determine the effectiveness of the compliance program. It will also consider sending self-review letters to a larger number of middle/lower risk groups which we have not selected for audit review to observe any changes on behaviour over a period of time. This will indicate the effectiveness of IRAS's educational as well as publicity approach in boosting voluntary compliance level.

Case study 13: Dealing with company directors for failure to file company accounts (Singapore)

Risk treatment

Background	IRAS noted that there was a high non-attendance rate among the small companies (turnover less than S$5 million dollars, usually with only 1 to 3 directors on board) when summonsed to court for non-submission of company Income Tax Returns and accounts. This prompted it to take enforcement action against company directors by issuing a notice to the them requesting a company's accounts, thus making the director personally liable for the submission of the

	document where non-compliance would result in the director being summonsed to court. When the directors responded with the accounts, IRAS got them to submit the company's outstanding returns at the same time.
Objectives	To improve filing compliance rate
Target group	The target group is the small companies with turnover less than S$5million
Risks treated	Failure to file on time
Key elements of risk treatment strategy	1) Empowerment in relevant Tax Act 2) Ensuring the company's directors information are up to date; 3) Frontline officers to deal with directors queries and getting them to file the outstanding returns together with the accounts; 4) Non-compliant directors need to be summons to court promptly.

Evaluation

Description of evaluation strategy approach	To evaluate the effectiveness, IRAS measured the number of cases that responded in terms of coming forward to voluntarily inform it of the company's dormant status as well as the court attendance rate.
Specialised support used	Data mining tool for case identification and monitoring.
Observed impacts of strategy	Number of cases compounded, court attendance rate and number of dormant companies coming forward voluntarily have increased/improved.
Other comments reflecting on the results achieved	IRAS has been able to flush out cases where business had ceased but directors fail to inform us. Lifting of corporate veil is necessary at times. As above, the directors came forward when they were made personally liable and realised that they would face a summons for non-compliance.

Case study 14: Higher risk new GST registrants (Singapore)

Risk treatment

Background	New registrants are not familiar with GST laws and tend to commit errors. (The risk treatment described has been integrated as a business-as-usual approach and is conducted yearly.)
Objectives	• To help traders with GST compliance at the onset to prevent downstream problems. • To create audit presence among new registrants.
Target group	• Businesses that registered within the last 15 months. • To audit new GST refunds claimed by such businesses.
Risks treated	Failure to correctly report liabilities
Key elements of risk treatment strategy	Strategy is to audit, make educational contacts and send targeted advisory letters. Specifically; • Risk rating the taxpayers to size up profile • Conduct telephone calls and interviews on their business model, how they prepared GST declaration and GST knowledge of preparers etc. • Audit on records, sales and purchases listing etc

Evaluation

Description of evaluation strategy approach	Results measured for evaluation purposes include 1) audit coverage; 2) the number of fraud cases detected; 3) audit recovery; 4) nature of error uncovered; and 5) strike rates. There is also a taxpayer survey on effectiveness and usefulness of the educational efforts (which is expected to be positive).
Specialised support	Taxpayer Survey on the new registrants to gauge their perception of our efforts and whether they find that help is readily available.
Observed impacts of strategy	These have been as follows: • Effective in detecting errors among the new registrants based on results collated for last 3 years. • Good gatekeeper in detecting fraud cases. • Consistent strike rate, nature of errors uncovered by team, staff perception of

	improved GST compliance among the new registrants and greater taxpayer awareness of common errors are good indicators of the effectiveness of our strategy.
Other comments reflecting on the results achieved etc	Taxpayers are engaged earlier than was the case prior to the strategy being adopted. However, IRAS acknowledges the need to do a targeted post-audit review of the same group that has been contacted to see if same errors are re-committed as part of a more complete evaluation of program effectiveness.

Case study 15: The 2004 anti-avoidance policy statement (re avoidance of national insurance contributions and PAYE income tax) (United Kingdom)

Risk treatment

Background	Until 2005 successive Governments were faced with increasingly more complex and contrived schemes to avoid national insurance contributions (NICs) & the collection of tax through the PAYE system on bonuses received by wealthy individuals. Up till then, each time one loophole was closed by the Government, avoidance activity would shift elsewhere.
Objectives	The Government in the 2004 Pre Budget Report decided it was time to take decisive action and ensure that all individuals would pay their fair share of tax and NICs however the reward is delivered. The aim of this announcement was to cause a permanent change in avoidance behaviour, bringing an end to remuneration-based tax avoidance.
Target group	Primarily wealthy individuals with high earnings. Often disproportionately found in the financial services sector.
Risks treated	Failure to correctly report liabilities
Key elements of risk treatment strategy	The 2004 Pre-Budget Report announced that all remuneration-based avoidance schemes which were known about would be closed down. Should further schemes be developed in future the Government would close them down also, where necessary from date of the statement - 2 December 2004. At that time the Government also introduced legislation to make employers/ users of NICs avoidance products disclose their schemes to HMRC.

Evaluation

Description of evaluation strategy approach	The evaluation comprised two main elements, both examining changes in avoidance behaviour in the 2005 and 2006 post-announcement years: • An analysis of individual tax payer data using econometric techniques (Difference-in-Differences) to estimate if there has been an increase in the total amount of tax paid as a proportion of their total income (their effective tax rate) or a change in the remuneration packages of avoiders that indicates a change in avoidance behaviour. • A detailed analysis of a small sample of 50 tax payers' tax-returns, examining the type of avoidance schemes that were in use before and after the announcement. Specifically, there were two treatment groups: A) Avoiders—individuals who were paid under PAYE by companies who were known by HMRC to provide their employees with remuneration –based avoidance schemes in the past; and B) A subset of avoiders—who also had positive dividend income. The evaluation compared the behaviour of these two groups against the behaviour of other taxpayers with similar incomes (who acted as a control group).
Specialised support	The internal HMRC evaluation was a contribution to HMRC's evidence-based policy-making, and was supported by external advice from the Centre for Employment Research at Westminster Business School.
Observed impacts of strategy	The evaluation, by focusing on employer-provided avoidance schemes, found robust evidence that the anti-avoidance legislation has raised additional revenue of £110 million in 2005 and £210 million in 2006. In 2004/05, the evidence suggests that the majority of the policy impact in the first (partial) year after the announcement operated through a reduction in avoidance using dividend-based schemes. For this group (category B) above), HMRC saw a large (and statistically significant) 5.6% point increase in their effective tax rate. In 2005/06 the evidence suggests that the effect of the policy on users of dividend-based schemes continued at a similar level. However the effect also widened to include other avoidance activity. The overall effect was that the wider target group (category A above) saw a large (and statistically significant) 2.8% point increase in their effective tax rate.
Other comments	Despite the retrospective element, the measure was not particularly controversial and was generally well received by the press and public. As well as being judged successful by HMRC's

reflecting on the results achieved & lessons learned	evaluation, comments issued by the press and by external experts also suggest the measure has been successful. Performing the evaluation in-house (but with support from external consultants) had several attractive features (including higher skills-transfer and skills-retention, and data-issues) compared to commissioning the evaluation externally.

Case study 16: Persistent Defaulters Letter Trial with VAT Debts (United Kingdom)

Risk treatment

Background	The primary purpose of the persistent defaulter letter trial was to test the effectiveness of the new DMB Persistent Debtor Letter in getting VAT persistent defaulters to pay their debt more quickly. Persistent Defaulters were taxpayers with an established history of default in the VAT system. The pilot was run over a period of five months.
Objectives	To collect VAT debt faster from taxpayers with a history of taking a long time to pay.
Target group	A persistent defaulter is defined as a trader who has just re-entered debt and who: • Has had 2 or more concurrent debt cases in the previous 12 months OR • Has had 2 or more surcharges and one or more concurrent debt cases in the previous 12 months; OR • Has been on a 15% surcharge and has had one or more concurrent debt cases in the previous 12 months Traders selected for the trial owed money of between £3,000 and £100,000. Traders also needed to pass a manual selection process to assess that they hadn't already called in or on specially agreed terms. Over three months, we also ran the trial in a subset of Debt Management units Liverpool and Halifax in the first month and Chesterfield and Southampton in the following four months. Over 500 taxpayers passed these criteria in the five months of the trial.
Risks treated	Failure to pay on time
Key elements of risk treatment strategy	Taxpayers were sent a letter explaining that HMRC had observed that they had a history of defaulting with them and were urged to call in and either pay what they owe, make a payment promise or time to pay agreement.

Evaluation

Description of evaluation strategy approach	HMRC identified that it wanted a specific volume of taxpayers to split 50/50 between trial and control samples. It then calculated how long the trial would have to run within the specified parameters. Samples were split randomly by HMRC's Knowledge and Intelligence (KAI) specialist group. Measures of success were faster debt payment and the taxpayer making contact with HMRC.
Specialised support	MS SQL Server and Excel were used.
Observed impacts of strategy	Significant increases were observed in taxpayers: 1) making contact with HMRC; 2) making payments within 14 and 28 days of the letterbox date; 3) clearing debts; and 4) making Payment Promises and Time to Pay Agreements
Other comments reflecting on the results achieved & lessons learned	There was unusually low level of background debt collection activity during the duration of the trial due to resources being dedicated to the Business Payment Support Service. Getting client commitment and agreement on the parameters and process of the trial would have meant fewer last minute changes at the implementation stage of the trial. The small scale trial run in the first month, although not planned, turned out to be a beneficial learning opportunity to help us refine the trial for the second month.

Case study 17: Tax avoidance arrangement/ scheme involving use of loss attributing qualifying companies (LAQCs) to fund the cost of taxpayers' private home expenses (New Zealand)

Risk treatment

Background	IRD was concerned about abuses of the tax law involving LAQCs. Specifically, it was concerned about people (i.e. shareholders in an LAQC) selling their own family home to the LAQC then renting the property back to themselves at a market rent. The LAQC then claims expenses relating to the house which would otherwise be personal expenses. Because

<table>
<tr><td></td><td>property outgoings (e.g. mortgage costs, rates) of the LAQC are higher than the rent received a loss arises which is then attributed to the LAQC shareholder(s) under the LAQC rules. Shareholders then offset this loss against their other income streams, effectively getting a deduction for part of their home expenses.</td></tr>
<tr><td>Objectives</td><td>To stop/ mitigate the abuse. Specifically to; 1) increase awareness in the general population; 2) increase awareness in the advisor population; 3) improve voluntary compliance; 4) provide further clarification of IRD’s position; 5) support the Revenue Alert.</td></tr>
<tr><td>Target group</td><td>An overall LAQC population was refined to a target group of 48,000 using specially devised screening criteria</td></tr>
<tr><td>Risks treated</td><td>Failure to correctly report liabilities</td></tr>
<tr><td>Key elements of risk treatment strategy</td><td>This entails: 1) issue of Revenue Alert which is a document from the Commissioner to the public, highlighting IR’s concerns on an issue and setting out IR’s position on that area of concern; 2) the testing and development of an information campaign through letters and pamphlets posted to the targeted population of 48,000, inviting voluntary disclosures with reduced shortfall penalties as an incentive (NB: This step was preceded by sample testing of letters sent to just under 2,000 entities/cases); 3) a publicity program on the IR website and in print media describing the actions to be taken; and 4) audits of cases resulting from further screening, after the information campaign and the opportunity to correct.</td></tr>
</table>

Evaluation

<table>
<tr><td>Description of evaluation strategy approach</td><td>LAQC measurements of impact include traditional revenue/treasury measures (i.e. basic hours expended and dollars received), behaviour, and awareness / sentiment / knowledge.

Revenue: Tracking the revenue gathered through voluntary disclosure and through Audit. Treasury reporting on hours and dollars.

Behaviour: Has the audiences behaviour been influenced by the campaign. Has the target group become more compliant in their LAQC. Impact on new LAQC registration volumes and LAQC deregistration volumes.

Awareness / Sentiment / Knowledge
• Blog analysis to determine if the campaign was having an impact by creating an increase in “buzz” and was it raising awareness in the digital arena,
• Sentiment filter to determine if the online response to the campaign was positive, neutral or negative,
• Media analysis to assess the interest generated by releases and articles including an online assessment,
• Measuring traffic volumes to and in the IR website from external search engines, the site’s internal search engine and from online ad’s to ascertain if taxpayers were seeking further knowledge.
• Two sets of qualitative interviews to examine impact & efficacy of educative efforts:
    o Community relations advisors - their perceptions on reception of the educative efforts; and
    o Influencers and intermediaries – their awareness of initiatives; and changes in their behaviour and/or the behaviour of their clients
• Text analysis of in-bound calling to identify impacts on behaviour and sources of information being acted upon</td></tr>
<tr><td>Use of specialised support</td><td>• Screening of taxpayer database to identify and refine target populations;
• Linking of External data sources (e.g. utilities, various service providers) to administrative data to further refine target lists;
• Google Analytics to track & measure website performance & user behaviour on the site;
• Internal Web traffic counts;
• Blog Analysis (Using a third party service called Buzzmetrics);
• APN (major media organisation in NZ);
• GIS (Geographical Information System).</td></tr>
<tr><td>Observed impacts of strategy</td><td>The key drivers of success for determining the effectiveness of the LAQC Risk Treatment to change compliance are:
1. Reduction/removal of loss claims,</td></tr>
</table>

	2. A reduction in the number of new entrants, 3. Volume of disclosures, $ gathered, 4. Increases in digital media online viewing 5. Increase in online discussion in special interest chat rooms. Measures for the above are achieved by undertaking a cyclic review of: • Changes in the risk population by re-running the original select, • The change in mix of Agent clients utilising LAQC structures, • Changes to the 'Nature of Business' field in LAQC registration forms, • Number of Voluntary Disclosures, • The background behind the taxpayers original motivation that recommended them into this structure; e.g., Advice from an Agent, recommendation from a friend-of-a-friend etc • Determine level of interest among current or prospective LAQC director's through the level of requests for Customer Insight (outbound contact) presentation's and seminars compared to a test group of LAQC director's we have not contacted. Measures to determine the effectiveness of our messaging to lift awareness include: • Counts of the number of LAQC related searches on the IR Website, • A suite of media measures, • Level of compliance outbound call contact and enquiry, • Site statistics (by subject area) (e.g. average time spent on LAQC pages, page views • Reference of Call Centre staff to LAQC in the IR Knowledge Base, • Monitoring topical discussions in the blogosphere and social networks, • Monitoring awareness levels via surveys; i.e., repeat Colmar Brunton survey or similar and compare against May 2008 baseline (generic compliance), together with other external sources of research; e.g., Commercial Bank Property Survey etc, • Monitoring the level of support from Accounting Firms and industry bodies such as the accountants' association, NZICA, including take-up on our Agent Answers, requests for presentations, response to feedback requests, utilisation of IR prepared material in their communications. **Behaviour** Definitive change in behaviour can be pinpointed and linked to campaign elements. • There has been a significant reduction in the registration of new LAQCs, • There has been a significant increase in the deregistration of existing LAQCs. Based on an average LAQC claim of $25k per LAQC and assuming that those who have changed to being compliant continue to be compliant the treatment is estimated to increase revenue by $638 million per year in future years. **Knowledge** • Traffic from external search engines looking for LAQC information increased 225% compared to the same time in the previous year. • Searches using the IRD website search engine increased by 23% compared to overall IRD site search traffic (both internal searches and external search engines) which increased by 16%.

Annex 2

References on evaluation

1) Australian Taxation Office's evaluation-related products

An overview of our methodology and its development: Outlines the ATO's compliance effectiveness methodology, its development, application and implementation.
See *Measuring compliance effectiveness – Our methodology NAT 72341 (PDF, 270KB)*

A guide to applying the methodology, for use in facilitated workshops: Guides workshop participants through the four phases of the ATO's compliance effectiveness methodology. See *Measuring compliance effectiveness – Applying our methodology NAT 72342 (PDF, 220KB)*

A guide to developing effective compliance strategies: Assists compliance staff in analysing compliance behaviour, and the various factors which can drive this behaviour, as they design and implement compliance strategies. The guidance information is complemented with a series of case studies.
Developing effective compliance strategies

A review of literature on measuring compliance effectiveness: Reviews Australian and international literature on measuring effectiveness with an emphasis on how the concepts have been used by revenue bodies. See *Literature Review - Measuring compliance effectiveness – 2007*

2) European Union

The European Commission's Fiscalis program has prepared practical guidance on risk management, including evaluation, as part of a Compliance Risk Management Guide for Tax Administrations, a revised edition of which was published in March 2010 —see http://ec.europa.eu/taxation_customs/resources/documents/common/publications/info_docs/taxation/risk_managt_guide_en.pdf

3) UK Treasury

THE GREEN BOOK: Appraisal and Evaluation in Central Government (UK Treasury): This guidance is designed to promote efficient policy development and resource allocation across government. See http://www.hm-treasury.gov.uk/d/green_book_complete.pdf?bcsi_scan_3CB14DF0471C3DC0=0&bcsi_scan_filename=green_book_complete.pdf

4) US Department of Health & Human Services, Center for Disease Control & Prevention

Introduction to Program Evaluation for Public Health Programs: A Self-Study Guide (August 2005); This document is a "how to" guide for planning and implementing evaluation activities. The manual is based on CDC's *Framework for Program Evaluation in Public Health*, and is intended to assist state, local, and community managers and staff of public health programs in planning, designing, implementing, and using the results of comprehensive evaluations in a practical way. See http://www.cdc.gov/GetSmart/program-planner/Introduction.pdf

5) OECD

Managing and improving compliance:
http://www.oecd.org/dataoecd/44/19/33818656.pdf?bcsi_scan_3CB14DF0471C3DC0=0&bcsi_scan_filename=33818656.pdf

Recent developments in compliance risk treatments:
http://www.oecd.org/dataoecd/36/34/42490764.pdf?bcsi_scan_3CB14DF0471C3DC0=0&bcsi_scan_filename=42490764.pdf

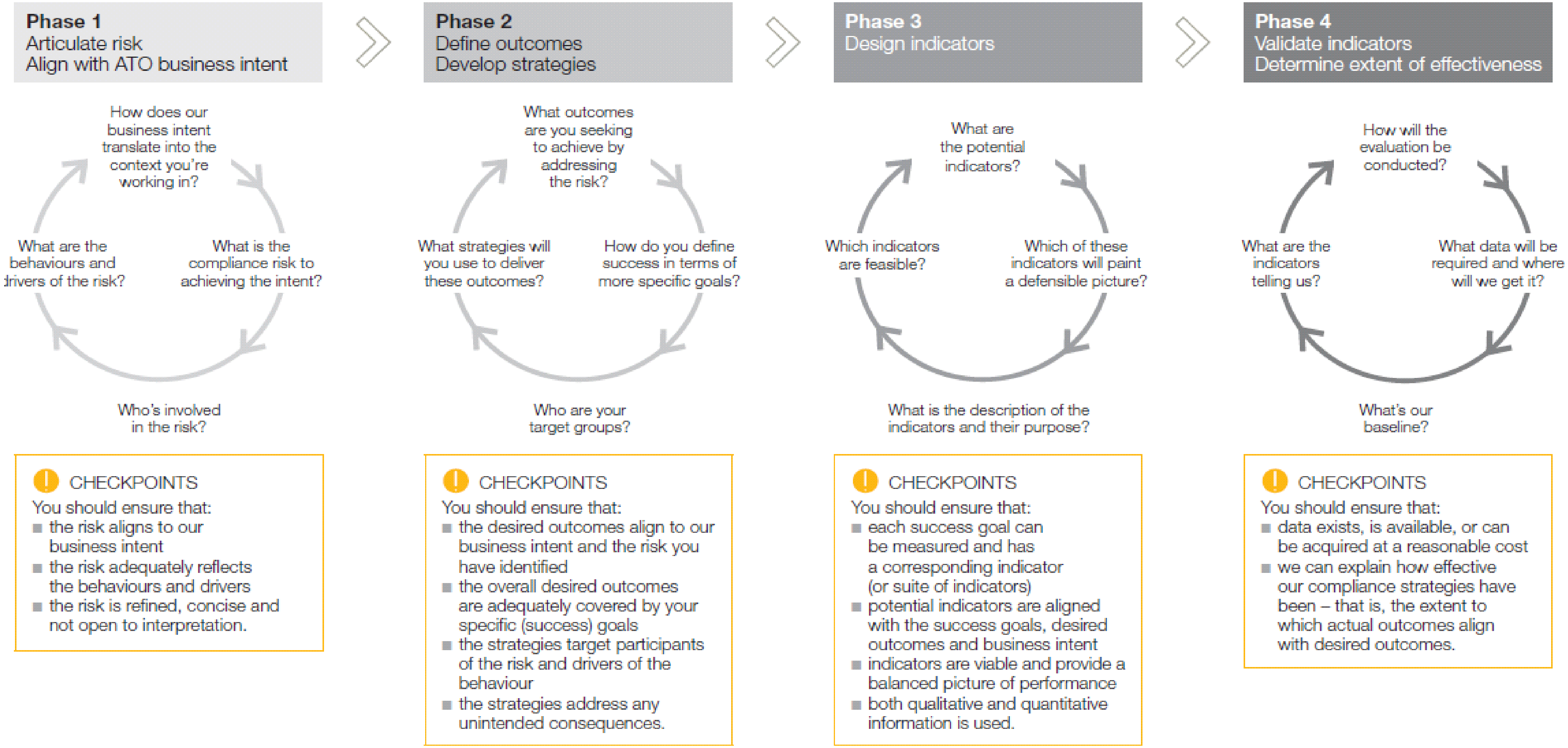
6) Other technical references

US General Audit Office, Designing Evaluations (1991): This methodology transfer paper addresses the logic of program evaluation designs. See http://www.gao.gov/special.pubs/10_1_4.pdf

US General Accounting Office, Program evaluation—an evaluation culture and collaborative partnerships help build agency capacity (May 2003): To assist agency efforts to provide credible information, GAO examined the experiences of five agencies that demonstrated evaluation capacity in their performance reports. See http://www.gao.gov/new.items/d03454.pdf

Annex 3

Figure 15. The ATO's compliance effectiveness methodology



Annex 4

Table 4. The SMART test elaborated

Test description	Example
Specific (S): Indicators should have a clear, unambiguous description so that everybody has a common understanding of the information the indicator is expected to deliver. Indicators should be expressed in a way that will help to identify whether compliance behaviour has changed as a result of the strategies employed. A simple way to do this is to describe the indicator as '*a change* or *a difference in the rate* or *ratio* or *number of …*' Comparing the compliance behaviour before you implemented your strategies to the behaviour afterwards allows you to understand whether there is a change in behaviour that correlates with your strategies'.	*A change in the percentage of income tax returns filed by the due date by sole traders with a turnover of less than $1 million.*
Measurable (M): Consideration needs to be given as to how changes from one period to the next are going to be measured. The following questions can be used in this context: • can the indicator be expressed in a numeric form? • will the data be compared against some form of benchmark? • is a calculation such as a ratio involved? What are the components of the calculation? • will the whole targeted population be measured or will the analysis be based on a sample?	A change in average rental income. This is calculated by dividing the total dollar value of rental income by total number of taxpayers reporting rental income.
Achievable (A): Indicators should be both practical and reasonable. The data for each indicator should be readily available or can be obtained at a reasonable cost.	A change in net tax payable to total income ratio. This indicator may be based on data that are readily available from a revenue body's database of tax return information
Relevant (R): Indicators should measure the expected or desired changes. They should also have a direct link with intended success goals and desired outcomes. An indicator should not be chosen simply because it is easy to measure. Efforts should be to ensure it contributes to the evidence being gathered to help understand the extent to which a revenue body's success goals have been achieved. What the indicator should show in the immediate and longer term needs to be understood. For example, a successful strategy may show a significant improvement in the short term, but one that levels off in the future – an immediate effect with a sustained change.	**Success goal:** Individual taxpayers voluntarily comply with their tax return filing obligations. **Indicator:** A change in the percentage of individual tax returns file by the due date. An increase in this indicator will provide evidence of improved voluntary compliance with filing obligations.
Timed (T): It is critical to establish whether the strategies employed affect voluntary compliance and community confidence over time. Indicators should be selected that will identify change and show progress over the immediate, intermediate and long term. Indicators also need to be based on data that can be produced regularly enough to track progress and quickly enough for it to be useful, with only a short time between the period the data covers and when it becomes available. Consideration should be given as to whether: 1) there is a time lag before the data becomes available; 2) the data is available on a regular or infrequent basis; and 3) the data is available for all of the relevant time period. There is often a time lag between the collection and availability of external data. You need to carefully examine your data sources to understand how timing issues will affect the usability of your indicator.	Change in the number of individual income tax returns filed compared to the number of people identified in the central statistical agency's *'labour force statistics'*. These statistics are generally presented as data which spans a calendar year rather than an income year. There is a notable time lag between collection and presentation. When comparing such data against tax records, ensure the correct months are represented in the data set for the income year rather than calendar year, and that they correspond to the year in which return filing behaviour is being assessed.

Source: ATO